THE UNIVERSITY
of EDINBURGH

Patterns of distribution of tree species in the neotropical lowland rainforest biome

By Julieth Serrano

THE UNIVERSITY *of* EDINBURGH

This thesis is submitted to the University of Edinburgh for the degree of Doctor of Philosophy

School of Biological Sciences

Institute of Molecular Plant Sciences

2017

Declaration

I hereby declare that the work contained in this thesis is my own, unless otherwise acknowledged and cited. This thesis has not in whole or in part been previously presented for any degree.

Julieth Serrano

2017

In the wet forest…"*I can only add raptures to the former raptures . . . each new valley is more beautiful than the last. It is not possible to give an adequate idea of the high feeling of wonder, admiration and devotion which fill the mind*"

Darwin, 1839

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

This thesis aims to explore distributional patterns of tree species in the neotropical lowland rain forest biome based on diversity analyses, dated phylogenies and species distribution models, using the family Sapotaceae as a case study. Sapotaceae is an abundant and diverse group in the neotropical lowland rain forest and its distributional patterns are representative of other tree clades in this biome. These characteristics make this family a good model to test ecological and biogeographic hypothesis in neotropical rain forests.

An analysis of beta-diversity measured by the number of shared species was used as a test of biotic homogeneity of Morrone's (2001) widely used system of neotropical biogeographic units. Biotic homogeneity was generally low, and Morrone's (2001) biogeographic regionalisation was found not to coincide with the distributional patterns of Sapotaceae species.

Divergence times of Sapotaceae species were estimated using a dated phylogeny based on DNA sequences of the nuclear ribosomal internal transcribed spacer (ITS) to explore the effects of Andean uplift, closure of the Isthmus of Panama and Pleistocene climatic changes on the evolutionary history of lowland rain forests in northern South America. The Andean uplift was found to have affected patterns of distribution by creating new habitats and altering hydrologic systems in northern South America, and in some cases by isolating lineages to the east and west of the Eastern Cordillera of the Andes. The closure of the Panama Isthmus and Pleistocene climatic changes do not seem to have strongly affected patterns of distribution or diversification in Sapotaceae. In general, the lack of congruent dates for many repeated biogeographic splits in the phylogeny (e.g., Amazon-Choco) suggests that idiosyncratic dispersal events have had a substantial effects on Sapotaceae's biogeography.

Finally, species distribution models generated for Sapotaceae in the Neotropics were used to identify areas of high predicted species richness in Colombia. The highest diversity of Sapotaceae species was predicted for the inter-Andean valleys and northern Amazon. These results were compared to the current system of Protected Areas in this country, demonstrating that areas of high conservation value based on predicted species richness have a low coverage of Protected Areas. Such gaps highlight the potential need for new systems for the delimitation of basic units for conservation at national levels in Colombia.

Acknowledgements

I would like to thank my supervisors James Richardson, Toby Pennington and Richard Milne. I would also like to thank Ivan Jimenez and Tiina Sarkinen. I am sincerely grateful for your professional and thoughtful advice, and the time and work you dedicated to this project.

I thank Michelle Hart, Laura Forest and Ruth Hollands for laboratory assistance. Thank you also to Peter Moonlight and the various working teams at the Royal Botanic Garden Edinburgh that kindly guided me at crucial times during my studies.

Terence D. Pennington and Felix Forest at the Royal Botanic Gardens, Kew are thanked for providing samples of Colombian Sapotaceae and permitting their extraction, respectively. Alain Franc supplied samples from French Guiana that were collected by Daniel Sabatier and Jean-François Molino all of whom are thanked for those contributions. Unpublished data has also been provided by Dayana Sanchez from the District University in Bogota, Colombia.

A georeferenced data set of Sapotaceae specimens covering previously unexplored collections was gathered thanks to Terence D. Pennington, Alan Elliott, Dairon Cardenas, Karol Sedano, Carolina and Rocio Cortez. Thank you for your contributions.

My work in the field was only possible thanks to Moises Rojas, Maria Fernanda Torres, Rocio Cortez, Fabio Avila, Doris Huertas, Andres Avella and Max Triana. Invaluable companionship, advice and support from local communities allowed me to visit stunning and plant diverse areas in Colombia, thanks to all.

For their advice during my fellowship at the Missouri Botanical Garden, thanks to Ivan Jimenez, Adam Smith, Sebastian Tello, Maribel Arenas and Isabel Loza.

I also thank the Bascom Fellowship from the Missouri Botanical Garden, the Darwin Trust and The Davis Fund from the University of Edinburgh, and the Royal Botanical Garden Edinburgh for the financial support.

To my friends and colleagues Karina, Maca, Mafe, Flavia, Pedro, Fernanda, Lucia, Ana, Emilio, Pakkapol, Yun Yu, Subhani, Surabhi, Mauricio, Andy, Andres, Vanessa, Roosvelt, Kyle, Javier and Eugenio, thanks for the laughs and patience. Thanks to Nico for the welcoming mornings in the reception at the Botanics. To Miguelito, Karina, Flavia, Vanessa and Fernanda thanks for our travels, I smile remembering them! I thank the RBGE Colombia team for their great work, little steps always count. To Karina, Moises, Mafe and Maca, thanks for your friendship and support during these years in Edinburgh. To Miguelito, thanks for being the best team.

For their unconditional love, support, friendship, for their trust and encouragement I thank my family, especially my mother, father, pi and Miguelito.

Introduction

The neotropical lowland rain forest biome hosts some of the most species-rich floras in the world. For instance, estimates of tree diversity in the largest area of lowland rain forest in South America, the Amazon, range from 6,727 to 16,000 (ter Steege *et al*., 2016; Cardoso *et al*., 2017). Other areas of rain forest can be found in the Neotropics from Central America to southern Brazil. Understanding how species diversity is distributed and has evolved in these rain forests is important fundamental knowledge, and also critical for conservation in times of unprecedented threats. This thesis focuses on one area - northern South America - whose rain forests, in communities like those occurring in the Amazon, the biogeographic Choco, Macarena, Catatumbo and inter-Andean valleys of Colombia, have been neglected in studies of species diversity, distribution and evolution of neotropical lowland rain forests (e.g., Hopkins, 2007).

The origins and distribution patterns of taxa in the neotropical lowland rain forest biome have been previously studied for its economic, biotic and conservation value. Such studies have for example used tools such as cladistics, phylogenetic analyses and expert knowledge on distributions to delineate units thought to host distinctive biotic communities that may share common evolutionary histories. These units have been aggregated into systems called biogeographic regionalisations and include the widely used terrestrial ecoregions by Olson et al., (2001), the zoographic regions by Holt *et al*., (2013) and the biogeographic units of Latin America and the Caribbean by Morrone (2001). Olson *et al*., (2001) and Holt *et al*'s., (2013) systems depict units based on the aggregation of taxa at a global scale, whereas Morrone's (2001) system delineates units based only in neotropical communities.

Biogeographic regionalisations in the Neotropics are important tools in conservation planning, but before they are adapted to national and local scenarios, properties of units within those systems (e.g., biotic homogeneity), and their correspondence to the distribution patterns of taxa representative of local and national ecosystems should be tested by data-driven studies. In the past such tests have been scarce, and national plans aiming to preserve natural resources have used global or continental biogeographic regionalisations to build local systems of biogeographic

units and to identify areas of conservation priority, lacking knowledge on the relevance of such systems at the country level (e.g., current units for conservation priority in Colombia).

The origins and distribution patterns of the neotropical flora, including the lowland rain forest biome, have not only been studied based on biogeographic units, other methods have used for instance dated phylogenies to explore the role of vicariance and dispersal on the evolution of representative groups in these areas (e.g., Winterton *et al*., 2014). There is evidence for the effects of mountain barriers on the isolation of populations in taxa such as *Cremastoperma* and *Dussia* (Pirie *et al*., 2006; Winterton *et al*., 2014). In other groups such as Protieae migration across barriers has been demostrated (Fine *et al*., 2014). How these findings relate to the evolutionary history of the lowland rain forest of northern South America is partially unknown as those studies have not included data from the diverse vegetation of communities like those occurring in the Colombian lowlands (e.g., Hopkins, 2007). Vicariance events in northern South America may have been caused by the Andean uplift (e.g., Pirie *et al*., 2006), the closure of Panama (Hoorn *et al*., 1995; Hoorn, 2010; Farris *et al*., 2011; Hoorn and Flantua 2015; Bacon *et al*., 2015), marine incursions (e.g. Antonelli *et al*., 2009) and the Pleistocene climatic changes (Haffer, 1969; Richardson *et al*., 2001 but see Colinvaux *et al*., 2001 and Whinnett *et al*., 2005).

Additional approaches for the study of the lowland rain forest biome in the Neotropics, have relied on species distribution modelling (e.g., Hopkins, 2007). Species Distribution Models (SDMs), based on environmental variables, use known occurrences to predict the potential distribution of taxa. They can be applied to various temporal and spatial scenarios (Franklin, 2013), if used to depict patterns of aggregation at present times within a geographic space delineated by national borders and by previously proposed biogeographic units, they could serve as a proxy to test the accuracy of those biogeographic units in depicting the distribution of representative taxa within national and local ecosystems.

Aims and structure of the thesis

In this dissertation, I aim to explore the distribution patterns of tree taxa in the neotropical lowland rain forest biome by using the plant family Sapotaceae as a model group. Sapotaceae is an abundant and species-rich family of trees that is an important component of neotropical rain forests. I also aim to relate distribution patterns of tree taxa in the Neotropics to conservation planning, at the national level, using Colombia as a case study.

To do this, in Chapter one, I explored the accuracy of Morrone's (2001) biogeographic units in depicting patterns of aggregation of neotropical tree species. For this, I tested biotic homogeneity within Morrone's (2001) units by measuring variation in species composition in Sapotaceae within and among biogeographic units. Variation in species composition was measured as the number of shared species, and analyses were performed by controlling for collection bias and distance decay. A recently curated data set of 22,917 specimens of accurately geo-referenced occurrences in Sapotaceae species was used for these analyses.

In Chapter two, I studied the effects of the Andean uplift, the formation of the Isthmus of Panama, and the Pleistocene glacial cycles on the temporal and spatial diversification of tree taxa in the lowland rain forest biome in northern South America. For this, I used a species level biogeographic reconstruction of the subfamily Chrysophylloideae, obtained from the internal transcribed spacer (ITS) region of nuclear ribosomal DNA, and I dated splits and radiation events in this subfamily. Chrysophylloideae is the most diverse Sapotaceae subfamily in northern South America. The molecular data set used in these analyses included records of previously under-explored (in terms of Sapotaceae species) Colombian lowland forests.

In Chapter three, I identified areas of high predicted species richness in the Neotropics, and focussed on Colombian ecosystems, I determined if these areas were represented in the current national system of Protected Areas and in the current set of areas of conservation priority used in this country. To determine areas of predicted species richness I used Species Distribution Models based on occurrence records of Sapotaceae species.

By integrating different approaches for the study of the distribution patterns of tree taxa in the Neotropics, including spatial, ecologic and historical biogeographic analyses, I aim to build an integrative picture of the evolution, but also of the current state of biotas within the neotropical lowland rain forest biome. Also, by exploring how the depiction of the distribution patterns of representative taxa has been used in applied conservation at the national level, I aim to contribute and link scientific studies to pragmatic conservation planning.

References

Antonelli, A., Nylander, J. a a, Persson, C., & Sanmartín, I. 2009. Tracing the impact of the Andean uplift on neotropical plant evolution. Proceedings of the National Academy of Sciences. 106: 9749-9754.

Bacon, C. D., Silvestro, D., Jaramillo C., Smith, B. T., Chakrabarty, P. & Antonelli, A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences. 112: 6110-6115.

Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., … Forzza, R. C. 2017. Amazon plant diversity revealed by a taxonomically verified species list. Proceedings of the National Academy of Sciences. 114: 10695-10700.

Colinvaux, P. A., Irion, G., Rasanen, M. E. & Bush, M. B. 2001. A paradigm to be discarded: Geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of amazonian speciation. Amazoniana. 16: 609-646.

Farris, D. W., Jaramillo, C., Bayona, G., Restrepo-Moreno, S. A., Montes, C., Cardona, A., … Valencia, V. 2011. Fracturing of the Panamanian Isthmus during initial collision with South America. Geology. 39: 1007-1010.

Fine, P. V. A., Zapata, F. & Daly, D. C. 2014. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). Evolution. 68: 1988-2004.

Franklin, J. 2013. Species distribution models in conservation biogeography: developments and challenges. Diversity and Distributions. 19: 1217-1223.

Haffer, J. 1969. Speciation in amazonian forest birds. Science. 165: 131-137.

Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., … Rahbek, C. 2013. An update of Wallace's zoogeographic regions of the world. Science. 339: 74-78.

Hoorn, C., Guerrero J., Sarmiento, G. A. & Lorente, M. A. 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. Geology. 23: 237-240.

Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., … Antonelli, A. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution and biodiversity. Science. 330: 927-931.

Hoorn, C. & Flantua, S. 2015. An early start for the Panama land bridge. Science. 348: 186-187.

Hopkins, M. J. G. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. Journal of Biogeography. 34: 1400-1411.

Morrone, J. 2001. Biogeografia de America Latina y el Caribe. First ed. Zaragoza: CYTED, ORCYT-UNESCO, Sociedad Entomológica Aragonesa (SEA).

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., … Kassem, K. R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience. 51: 933-938.

Pirie, M. D., Chatrou, L. W., Mols, J. B., Erkens, R. H. J. & Oosterhof, J. 2006. 'Andean-centred' genera in the short-branch clade of Annonaceae: Testing biogeographic hypotheses using phylogeny reconstruction and molecular dating. Journal of Biogeography. 33: 31-46.

Richardson, J. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science. 293: 2242-2245.

ter Steege, H., Vaessen, R. W., Cárdenas-López, D., Sabatier, D., Antonelli, A., de Oliveira, S. M., … Salomão, R. P. 2016. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. Scientific Reports. 6: 29549.

Whinnett, A., Zimmermann, M., Willmott, K. R., Herrera, N., Mallarino, R., Simpson, … Mallet, J. 2005. Strikingly variable divergence times inferred across an Amazonian butterfly 'suture zone'. Proceedings of the Royal Society, London, B. 272: 2525-2533.

Winterton, C., Richardson, J. E., Hollingsworth, M., Clark, A., Zamora, N., & Pennington, R. T. 2014. "Historical biogeography of the neotropical legume genus Dussia: the Andes, the Panama Isthmus and the Chocó," in Paleobotany and Biogeography: A Festschrift for Alan Graham in his 80th Year (Eds.) Stevens, W. D., Montiel, O. M. & Raven, P. St Louis, MO: Missouri Botanical Press. 389- 404.

Lay Summary

The neotropical lowland rain forest biome hosts some of the most species-rich floras in the world. For instance, estimates of tree diversity in the largest area of lowland rain forest in South America, the Amazon, range from 6,727 to 16,000. Other areas of rain forest can be found in the Neotropics from Central America to southern Brazil. Understanding how species diversity is distributed and has evolved in these rain forests is important, fundamental knowledge, and also critical for conservation in times of unprecedented threats. This thesis focuses on one area - northern South America - whose rain forests, in communities like those occurring in the Amazon, the biogeographic Choco, Macarena, Catatumbo and inter-Andean valleys of Colombia, have been neglected in studies of species diversity, distribution and evolution. Here, I explore distributional patterns of tree species in the neotropical lowland rain forest biome based on diversity, spatial and DNA analyses, using the plant family Sapotaceae as a case study. Sapotaceae is an abundant and diverse group in the neotropical lowland rain forest and its distributional patterns are representative of other tree species in this biome. These characteristics make this family a good model to test ecological and biogeographic hypothesis in neotropical rain forests.

Chapter 1. Biotic Homogenity of Biogeographic Units in the Neotropics: A Test with Sapotaceae

1.1 Introduction

A central tenet of biogeography is that organisms are distributed across geographic space in a non-random fashion, forming spatial aggregations of endemic taxa with overlapping distributions, a phenomenon known as provincialism (Lomolino *et al*., 2010). At least since the nineteenth century provincialism has been described by drawing divisions on Earth that reflect patterns of biotic similarity (e.g., De Candolle, 1820; Sclater, 1858; Wallace, 1876), thus delimiting areas heralded as "biogeographic units". These units are often regarded as parts of hierarchical systems of nested areas (e.g. McLaughlin, 1992; Cracraft, 1994; Morrone, 2001; Kreft and Jetz, 2010; Holt *et al*., 2013, but see Stoddart, 1992) known as "biogeographic regionalizations". Realms or regions are the largest areas in these systems, frequently delimited according to the distribution of higher taxa, such as families and orders. Lower in the biogeographic hierarchy are increasingly smaller areas, for instance subregions and provinces, delimited according to the distribution of taxa at increasingly lower taxonomic ranks, including genera and species. In this way, subregions are nested within realms and provinces within subregions. These proposed biogeographic units, as well as their relationships in terms of biotic similarity, play significant roles in attempts to uncover the history and current spatial structure of life on earth (Lomolino *et al*., 2010), and in plans for the conservation and management of biological diversity (Whittaker *et al*., 2005; Ladle and Whitaker, 2011).

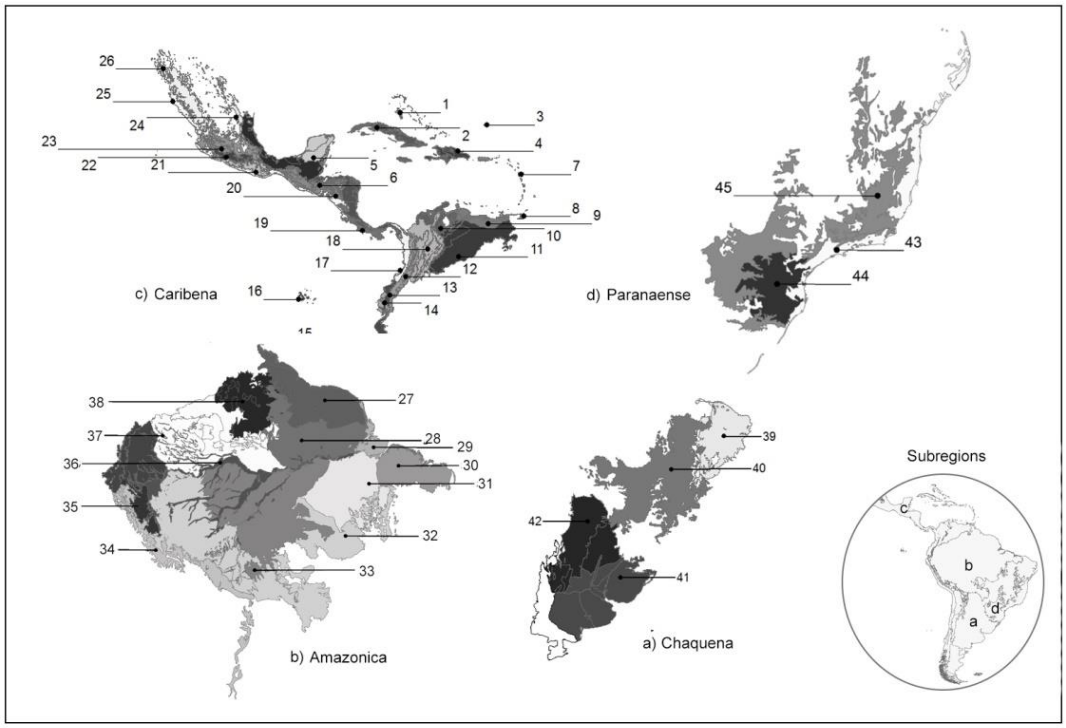
Despite its significance for basic science and biodiversity conservation, delineation of biogeographic units in tropical regions of Earth remains uncertain (Mackey *et al.,* 2008, Escalante, 2009, Kreft and Jetz, 2010). In no small part this uncertainty stems from the fact that many extant species have not been described (the Linnaean shortfall, Essl *et al*., 2013) and that the geographic distribution of described species is poorly known (the Wallacean shortfall, Sheth *et al*., 2012). The Linnaean and Wallacean shortfalls are particularly important impediments for the accurate

delineation of units at the lower ranks of the hierarchies proposed by biogeographic regionalizations. These units cover relatively small areas, and are based on the geographic distribution of taxa at low taxonomic ranks, often species. Therefore, their delimitation requires data with high spatial and taxonomic resolution. Because availability of these data may often be limited, proposed biogeographic regionalizations may best be regarded as working hypotheses that yield testable predictions (Whittaker *et al*., 2005; Mackey *et al*., 2008). As data with high spatial and taxonomic resolution become increasingly available, further progress can be made by testing these predictions, thus determining the extent to which putative biogeographic units constitute accurate portrayals of provincialism.

A key prediction implicit in proposed biogeographic regionalizations is the biotic homogeneity of putative biogeographic units. In particular, if putative biogeographic units accurately describe the spatial aggregation of endemic taxa (i.e. provincialism), then biotas should be more homogeneous within than among units (Stoddart, 1992; Lomolino *et al.*, 2006, page 342; Kreft and Jetz, 2010). In other words, change in taxon composition between sites within a given biogeographic unit should occur at lower rates than change in taxon composition between sites located in different biogeographic units. Several methods commonly used to propose biogeographic units at any level in the hierarchy (e.g. subregions or provinces), are designed to identify areas that contain distinctive endemic taxa (e.g., Szumik and Goloboff, 2004; Morrone, 2014). But any given area (even a randomly selected area) may contain distinctive endemic taxa and, nonetheless, lack spatially homogeneous biota relative to other areas (Lomolino *et al*., 2006, page 342). Other proposals of biogeographic units are based on analyses that cluster sites according to their biotic similarity (e.g., Kreft and Jetz, 2010; Mouillot *et al*., 2013; Oliveiro *et al*., 2013, Holt *et al*., 2013; Vilhena and Antonelli, 2015). However, biotic similarity among sites can be largely driven by geographic distance (Soinien *et al.,* 2007), and thus putative biogeographic units identified using this kind of criterion may not reflect true discontinuities but arbitrary division of a gentle gradient of taxon turnover across geographic space (Magnusson, 2004; Fortin and Dale, 2005; page 180). Therefore, empirical tests of the spatial discontinuities in taxon turnover predicted by proposed biogeographic units should ideally control for distance decay of biotic similarity

(Stuart et al. 2012). Such tests seem scarce (Lomolino *et al.*, 2006, but see Stuart *et al*., 2012); and yet they are required to examine the merit of putative biogeographic units, again, even when such units are known to host distinctive endemic taxa.

Here we view the neotropical biogeographic units proposed by Morrone (2001) as working hypotheses, and focus on testing if they are homogeneous in terms of one component of the biota: plant species in the family Sapotaceae. These biogeographic units, delineated according to the geographic distribution of vascular plants, insects and birds, hierarchically divide the Neotropical region into provinces nested within subregions (Fig. 1.1). This is the most comprehensive biogeographic regionalization currently available for the Neotropics. It is based on various kinds of analysis, including parsimony analysis of endemicity, and approaches in cladistic biogeography and panbiogeography. We are unaware of any test of the biotic homogeneity of any of the biogeographic units proposed by Morrone (2001). Here we address this gap using a recently assembled dataset, with relatively high spatial and taxonomic resolution, on the distribution of Sapotaceae species, an important component of the neotropical regional flora in terms of diversity and abundance (Pennington, 1990, 2007; Burnham and Johnson, 2004; Bartish *et al.*, 2011). Specifically, we tested if variation in species composition (beta-diversity) within biogeographic units was lower than across biogeographic units, while controlling for potential confounding effects of geographic distance and heterogeneous botanical sampling effort.



NEOTROPICAL REGION

Caribena subregion

1. Bahamas
2. Cuba
3. Puerto Rico
4. Espanola
5. Peninsula de Yucatan
6. Chiapas
7. Antillas Menores
8. Trinidad Tobago
9. Costa Venezolana
10. Maracaibo
11. Llanos Venezolanos
12. Cauca
13. Occcidente de Ecuador
14. Ecuador Arido
15. Tumbes Piura
16. Islas Galapagos
17. Choco
18. Magdalena
19. Occidente del Istmo de Panama
20. Oriente de America Central
21. Sierra Madre del Sur
22. Depresion de Balsas
23. Eje Volcanico Transmexicano
24. Sierra Madre oriental
25. Costa Pacifica Mexicana
26. Sierra Madre occidental

Amazonica subregion

27. Guyana Humeda
28. Roraima
29. Amapa
30. Para
31. Tapajos Xingu
32. Pantanal
33. Madeira
34. Yungas
35. Napo
36. Varzea
37. Imeri
38. Guyana

Chaquena subregion

39. Caatinga
40. Cerrado
41. Pampa
42. Chaco

Paranaense subregion

43. Bosque Atlantico Brasileno
44. Bosque Araucaria Angustifolia
45. Bosque Paraense

Figure 1.1. Neotropical biogeographic units proposed by Morrone (2001).
These were defined using a variety of methods including panbiogeography, cladistic biogeography and areas of endemism analyses. The proposed biogeographic units form a hierarchy in which regions are divided into subregions and subregions into provinces (from 1 to 45 in the legend).

1.2 Methods

1.2.1 Study group

Sapotaceae, currently classified within Ericales and comprising 53 genera and more than 1100 species (Govaerts *et al*., 2001; Pennington, 1991, 2007), is a group of trees and a few shrubs predominantly distributed in the tropics, and particularly diverse in Africa, Asia and South America in lowland and lower montane rain forest (Pennington, 1990, 1991; Swenson and Anderberg, 2005; Swenson *et al*., 2008).

The distributional range of the family in the American continent extends from the southern United States to northern Chile (Pennington, 1990, 2007). About 450 species of Sapotaceae are found within the Neotropics. They predominantly occur in forests below 1,000 m of elevation, reaching heights of 40-45 meters as canopy trees. A few species occur at higher altitude, including *Chrysophyllum lanatum* and *Pouteria lucuma*, both known from localities at 3,000 m of elevation. Sapotaceae is the second most abundant plant family in forests such as the Amazon (Ter Steege *et al.,* 2013), and it is indeed an important component of neotropical lowland rain forests, in terms of numbers of species and individuals (Pennington, 1990, 2007; Burnham and Johnson, 2004; Bartish *et al.*, 2011). Therefore, they are a useful model system for biogeographic studies in the Neotropics.

1.2.2 Species occurrence data

Occurence records of neotropical Sapotaceae species were compiled from the Herbario Nacional, Herbario Forestal, Herbario Amazónico Colombiano, Herbario de la Universidad del Valle, Herbario "Choco" in Colombia, and the PADME, GBIF (Global Biodiveristy Information Facility) and TROPICOS® databases. Duplicate records and records with ambiguous or tentative species level determination were excluded from the dataset. Specimen records missing geographic coordinates for the collection locality were geo-referenced at a spatial resolution of 1-degree, whenever enough information was available. The final dataset comprised 22,917 records representing > 90% of the estimated number of Sapotaceae species occurring in the Neotropics. We used this dataset to estimate the occurrence of Sapotaceae species in sampling units defined as 1-degree cells overlaid on the Neotropics (Fig. 1.2).

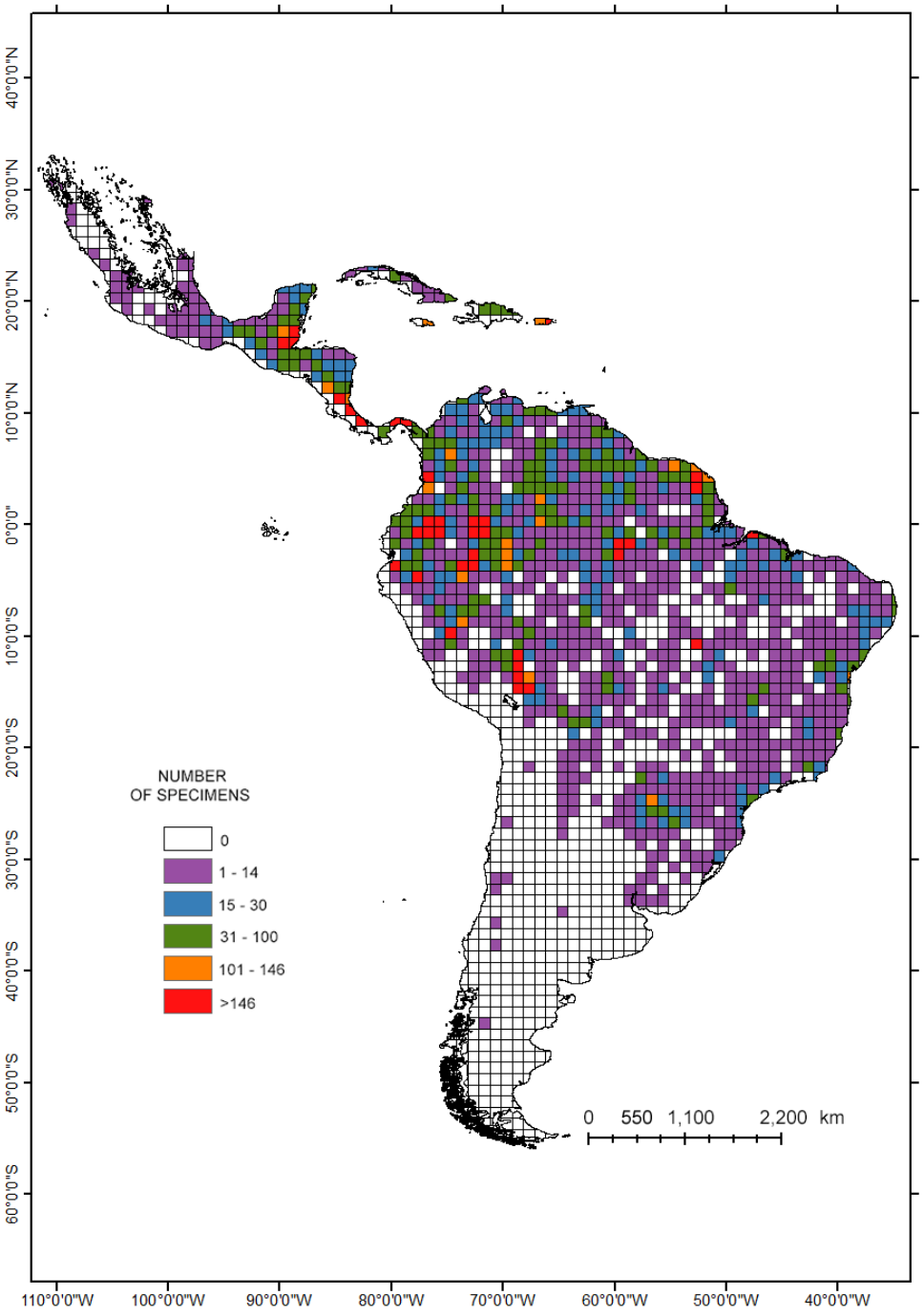


Figure 1.2. Geographic distribution of all 1-degree square cells with Sapotaceae specimen records.
Collection events for Sapotaceae specimens were overlaid in all sample units (1-degree square cells) in the Neotropical Region *sensu* Morrone (2001). An Albers conic equal-area projection was used. Collecting efforts have been concentrated in areas like the Amazonian forests and do not equally represent other areas in the known distributional range of Sapotaceae. Units in red represent the highest values in collection density, units in purple represent the lowest values in collection density.

1.2.3 Boundaries of putative biogeographic units

We divided the Neotropics into biogeographic subregions and provinces *sensu* Morrone (2001). To represent biogeographic divisions in a spatially consistent fashion, we aggregated polygons of the shapefile of world terrestrial ecoregions (Morrone, 2001; Olson *et al*., 2001; Olson and Dinerstein, 2002) into areas corresponding to Morrone's subregions and provinces using explicit synonymy between ecoregions and biogeographic units, and then assigning sampling units defined as 1-degree cells to the latter (Fig. 1.1).

1.2.4 Testing for biotic homogeneity while controlling for geographic distance and sampling effort

If putative biogeographic units are biotically homogeneous, then variation in species composition (i.e. beta-diversity, Anderson *et al.,* 2011) between sites within a biogeographic unit should be lower than that between sites located in different biogeographic units. We tested this prediction by developing a novel framework for comparing beta-diversity of Sapotaceae species between 1-degree cells located within a biogeographic unit to that between 1-degree cells located in different biogeographic units. We performed these comparisons conditional on geographic distance, to control for distance decay of biotic similarity (see introduction). In this way, we tested if beta-diversity of Sapotaceae species within biogeographic units was lower than beta-diversity of Sapotaceae species across biogeographic units, after correcting for the effect of geographic distance on species turnover. The biogeographic units proposed by Morrone (2001) are arranged in a hierarchy in which provinces are nested within subregions and subregions are nested within the Neotropical region (Fig. 1.1). Accordingly, we conducted nested comparisons of beta-diversity. In other words, for each pair of provinces within each subregion we tested if beta-diversity was lower within than across the provinces, and for each pair of subregions within the Neotropical region we tested if beta-diversity was lower within than between subregions.

We measured beta-diversity of Sapotaceae species between sampling units (i.e., 1-degree cells) as the number of shared species. This measure is influenced by differences in the number of observed species between 1-degree cells, which may stem from differences in botanical collecting effort. We therefore used a null model that controls for differences in the number of species between sampling units (Raup and

Crick, 1979; Chase and Myers, 2011). In this null model the assemblage of species in any given 1-degree cell was obtained by randomly sampling a "regional" species pool until the observed number of species in the 1-degree cell was matched. The "regional" species pool was defined according to the hierarchical structure of the biogeographic units proposed by Morrone (2001). Thus, when comparing beta-diversity within and across a pair of subregions belonging to the Neotropical region, the "regional" species pool was defined as the set of species known from the Neotropical region. Likewise, when comparing beta-diversity within and across a pair of provinces belonging to a particular subregion, the "regional" species pool was defined as the set of species known from the subregion containing the pair of provinces. For all species in the "regional" species pool, the probability of being part of the (null) species assemblage in any given 1-degree cell was proportional to the occupancy of that species in the "regional" species pool. In the case of comparing beta-diversity within and across a pair of subregions belonging to the Neotropical region, occupancy was the proportion of 1-degree cells occupied by the species across the Neotropical region. Likewise, when comparing beta-diversity within and across a pair of provinces belonging to a particular subregion, occupancy was the proportion of 1-degree cells occupied by the species across the subregion. For each pair of 1-degree cells we ran 1,000 iterations of the null model to generate a null distribution of the number of shared species, and then calculated the expected value (SSexp) and standard deviation (SSsd) of that null distribution. Using the observed (or actual) number of shared species (SSobs) we calculated beta-diversity for each pair of 1-degree cells as standardized effect size = (SSobs – SSexp) / SSsd. Note that this standardized effect size is inversely related to beta-diversity.

To test the prediction that, after correcting for geographic distance, beta-diversity within a biogeographic unit was lower than beta-diversity between biogeographic units, we used the following model of distance matrix regression (Legendre and Legendre, 1988) for every pair of provinces within each subregion, and for every pair of subregions within the Neotropical region:

$$(1),\ SES_{ij} = a_o + a_1 \cdot Z1_{ij} + a_2 \cdot Z2_{ij} + a_3 \cdot D_{ij} + a_4 \cdot D_{ij} \cdot Z1_{ij} + a_5 \cdot D_{ij} \cdot Z2_{ij}$$

where the response variable SES_{ij} is the standardized effect size (i.e. (SSobs – SSexp) / SSsd) for pairs of degree cells (i and j), $Z1_{ij}$ is a dummy variable with a value of 1 if both degree cells (i and j) are in biogeographic unit 1 (a province or subregion) and zero otherwise, $Z2_{ij}$ is a dummy variable with a value of 1 if both degree cells (i and j) are in biogeographic unit 2 and zero otherwise, and D_{ij} is geographic distance between degree cells i and j (measured as great circle distance). Finally, terms a_0 through a_5 in the right hand side of equation 1 are regression coefficients. Empirical support for the prediction (that beta-diversity within a biogeographic unit is lower than beta-diversity between biogeographic units) requires a_1 and a_2 to be statistically significant and positive, so that the regression lines for pairs of degree cells within biogeographic units would have a higher intercept than pairs of degree cells located in different biogeographic units. It also requires that these differences in intercepts do not fade away with distance (Fig. 1.3).

To test the statistical significance of the regression coefficients in equation 1 we created null distributions for each coefficient by permuting at random the rows of the response matrix and the corresponding columns, following the procedure described in Legendre and Legendre (1998). For every pair of provinces within each subregion, and for every pair of subregions within the Neotropical region, we simplified the regression model (equation 1) by excluding regression coefficients that were not statistically significant, following model simplification procedures in Crawley (2002). We examined regression coefficients and respective regression lines to determine if there was empirical support for the prediction that variation in species composition between sites within a biogeographic unit should be lower than that between sites located in different biogeographic units. In particular, for every pair of biogeographic units (provinces within each subregion or subregions within the Neotropical region) we determined if the mean number of shared species (measured as standardized effect size) for any given geographic distance was higher between 1-degree cells located in the same biogeographic unit than between 1-degree cells located in different biogeographic units (Fig. 1.3).

All analyses were performed in R 3.3.2. and maps and figures were produced using ArcMAP 10.1 and R 3.3.2.

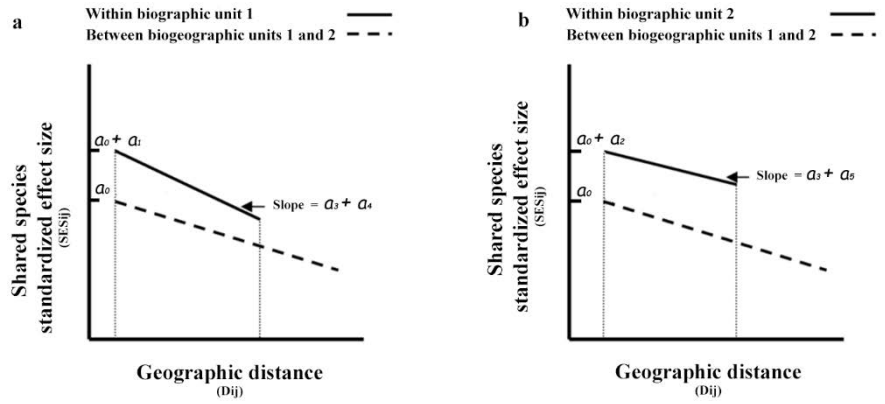


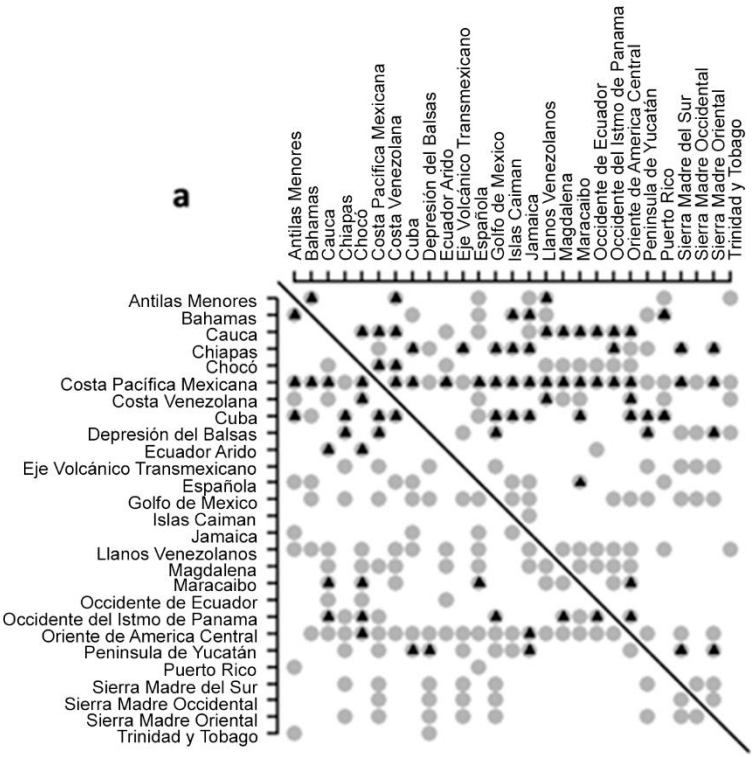
Figure 1.3. Predicted matrix regression results if putative biogeographic units accurately describe provincialism.
The mean number of shared species for any given geographic distance is described by regression lines according to equation 1, and should be higher for pairs of 1-degree cells located in the same biogeographic unit than for pairs of 1-degree cells located in different biogeographic units. Comparison of shared number of species was performed conditional on geographic distance, and thus only within the range of geographic distance shown by the vertical dotted lines in panels a and b. If coefficient a_1 in equation 1 is statistically significant and positive, then the regression line for pairs of degree cells within biogeographic unit 1 would have a higher intercept than pairs of degree cells located in different biogeographic units (a). Likewise, if coefficient a_2 in equation 1 is statistically significant and positive, then the regression line for pairs of degree cells within biogeographic unit 2 would have a higher intercept than pairs of degree cells located in different biogeographic units (b). Note that, potentially, results can differ between panels a and b: the number of shared species may be higher between 1-degree cells located in biogeographic region 1 than between 1-degree cells located in different biogeographic regions, even if the number of shared species is not higher between 1-degree cells located in biogeographic region 2 than between 1-degree cells located in different biogeographic regions.

1.3 Results

In the biogeographic regionalization proposed by Morrone (2001), the largest biogeographic units within the Neotropical region are the subregions Caribeña, Amazonica, Chaqueña and Paranaense (Fig. 1.1). Beta-diversity between 1-degree cells (sampling units) located in the same subregion was not generally lower than beta-diversity between 1-degree cells located in different subregions (Appendix 1.1). Thus, our analyses did not support the idea that the four neotropical subregions proposed by Morrone (2001) are biotically homogeneous with respect to each other in terms of Sapotaceae species.

According to Morrone (2001), each of the four subregions in the Neotropical region is divided into provinces. The northernmost of these subregions, the Caribeña subregion, is divided into 29 provinces (Fig. 1.1). Ideally, the biotic homogeneity of

each of these provinces would be examined with respect to all remaining 28 provinces in the subregion, by carrying out 28 pairwise comparisons between provinces to test if beta-diversity between 1-degree cells located in the same province is lower than beta-diversity between 1-degree cells located in different provinces. However, this was not possible because we had data for only 27 of the provinces. Also, the geographic distances separating 1-degree cells located in the same province did not always overlap the geographic distances separating 1-degree cells located in different provinces (Fig. 1.4), thus preventing comparison of beta-diversity conditional on geographic distance (Fig. 1.3). Nonetheless, tests of biotic homogeneity were possible with respect to nine or more provinces (mean number of pairwise comparisons = 9.37, standard deviation = 5.77, minimum = 0, maximum = 26). These tests showed that none of the provinces of the Caribeña subregion was biotically homogeneous with respect to all other provinces (Fig. 1.4). However, Cuba was homogeneous in 84% of the pairwise comparisons performed, Cauca in 75% and Costa Pacifica Mexicana in 73% (Figs. 1.4-1.5).



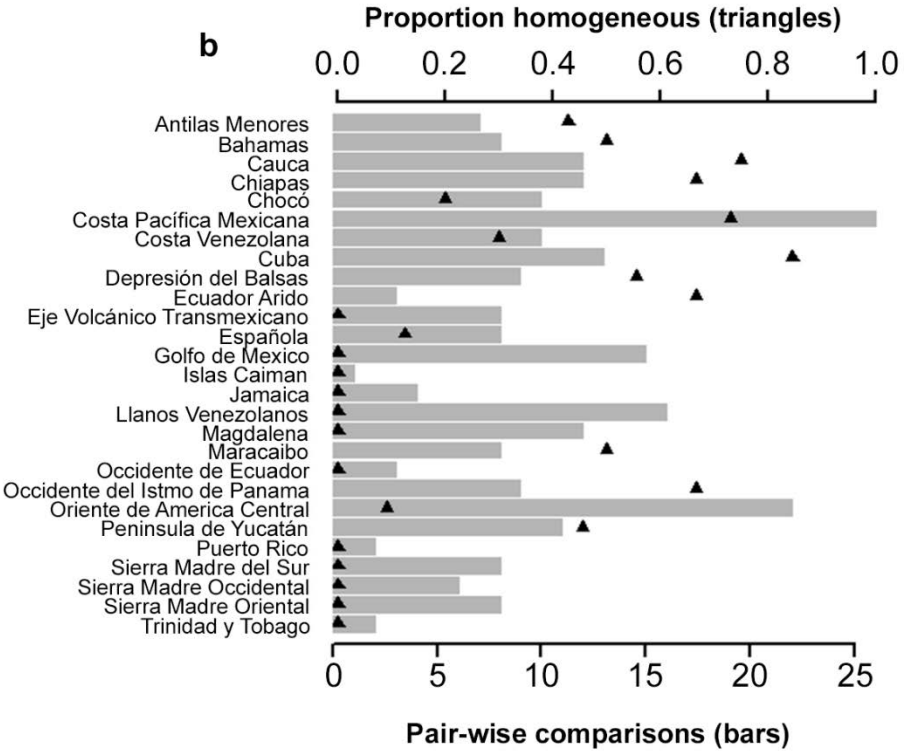


Figure 1.4. Summary of results from the test of biotic homogeneity for provinces in the subregion Caribena.
a) Results for pairs of provinces in the subregion Caribena. Each row shows the results of tests of biotic homogeneity of a particular province (i.e., the province after which the row is labeled) relative to each of the provinces in the columns. Gray dots indicate that there was overlap between the geographic distances separating 1-degree cells located within the (row) province and the geographic distances separating 1-degree cells located in different provinces. Thus gray dots show the pairs of provinces for which the test of biotic homogeneity could be (and was) performed. Triangles indicate that beta-diversity within the (row) province was lower than beta-diversity between the pair of provinces, thus empirically supporting the prediction of biotic homogeneity. b) Summary of results for each province, showing the number of pairwise comparisons for which the test of biotic homogeneity could be (and was) performed (gray bars, lower axis) and the number of pairwise comparisons for which the prediction of biotic homogeneity was supported (triangles, upper axis).

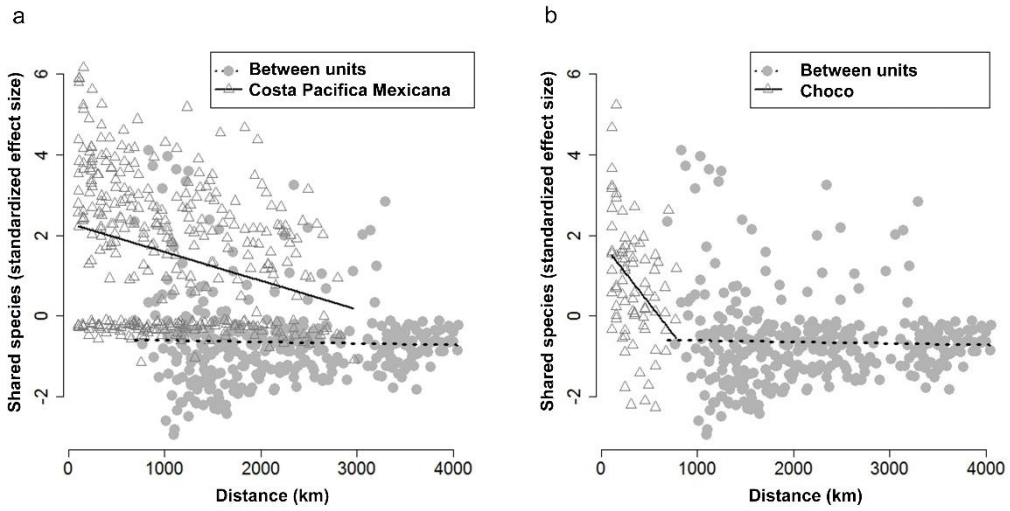


Figure 1.5. Comparison of beta-diversity within and between two provinces of subregion Caribena: Costa Pacifica Mexicana and Choco.
Beta-diversity within each of these provinces was lower than beta-diversity between them, thus empirically supporting the prediction of biotic homogeneity. a) Comparison of beta-diversity within province Costa Pacifica Mexicana to beta-diversity between provinces Costa Pacifica Mexicana and Choco. b) Comparison of beta-diversity within province Choco to beta-diversity between provinces Choco and Costa Pacifica Mexicana. Symbols represent pairs of 1-degree cells, and lines are matrix regression lines. Note that the regression lines for within province beta-diversity are higher than that for between province beta-diversity, thus supporting the prediction that variation in species composition between sites within a biogeographic unit should be lower than that between sites located in different biogeographic units.

At the centre of the Neotropical region is the subregion Amazonica, divided into 13 provinces (Fig. 1.1). For each of these provinces tests of biotic homogeneity were possible with respect to at least five other provinces within the subregion (mean number of pairwise comparisons = 8.92, standard deviation = 2.87, minimum = 5, maximum =12). None of the provinces of the subregion Amazonica was biotically homogeneous with respect to all other provinces (Fig. 1.6). However, the Varzea province was homogeneous in 75% of all possible pairwise comparisons (Fig. 1.6).

The two remaining neotropical subregions are Chaqueña and Paranaense, divided into four and three provinces, respectively (Fig. 1.1). For each of these provinces tests of biotic homogeneity were possible with respect to at least one other province within the respective subregion (Chaqueña: mean number of pairwise comparisons = 2, standard deviation = 0.82, minimum = 1, maximum = 3; Paranaense: mean number of pairwise comparisons = 2, standard deviation = 0, minimum = 2, maximum =2). Within the Chaquena, Chaco showed support in 100%, Cerrado in

67%, and Pampa in 50% of all possible pairwise comparisons (Appendix 1.1). In case of the Paranaense subregion Bosque Araucaria Angustifolia and Bosque Paranaense were biotically homogeneous relative to Bosque Atlantico Brasilero (50% of the two possible pairwise comparisons). Bosque Atlantico Brasilero was not biotically homogeneous compared to any other province (Appendix 1.1).

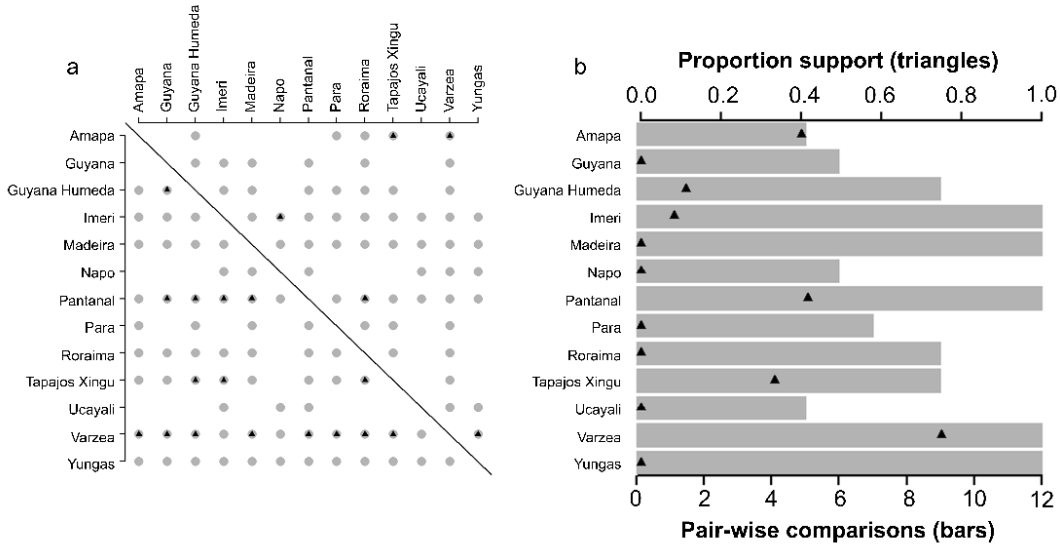


Figure 1.6. Summary of results from the test of biotic homogeneity for provinces in the subregion Amazonica.
a) Results for pairs of provinces in the subregion Amazonica. Each row shows the results of tests of biotic homogeneity for a single province (i.e., the province after which the row is labelled) relative to each of the provinces in the columns. Gray dots indicate that there was overlap between the geographic distances separating 1-degree cells located within the (row) province and the geographic distances separating 1-degree cells located in different provinces. Thus gray dots show the pairs of provinces for which the test of biotic homogeneity could be (and was) performed. Triangles indicate that beta-diversity within the (row) province is lower than beta-diversity between the pair of provinces, thus empirically supporting the prediction of biotic homogeneity. b) Summary of results for each province, showing the number of pairwise comparisons for which the test of biotic homogeneity could be (and was) performed (gray bars, lower axis) and the number of pairwise comparisons for which the prediction of biotic homogeneity was supported (triangles, upper axis).

In 136 out of 395 total pairwise comparisons performed, including subregions and provinces, beta-diversity between 1-degree cells (sampling units) located in the same biogeographic unit was lower than beta-diversity between 1-degree cells located in different biogeographic units at relatively short geographic distances. However, these differences did not hold at larger geographic distances (e.g., Fig. 1.7).

Specifically, in seven out of 12 pairwise comparisons within the Neotropical region beta-diversity was lower within the same subregion than beta-diversity between subregions only at relatively short distances (Appendix 1.2). This pattern was more frequent in the Caribena and Paranaense units. In Caribena three out of three and in Paranaense two out three total pairwise comparisons showed biotic homogeneity at short geographic distances (Appendix 1.2).

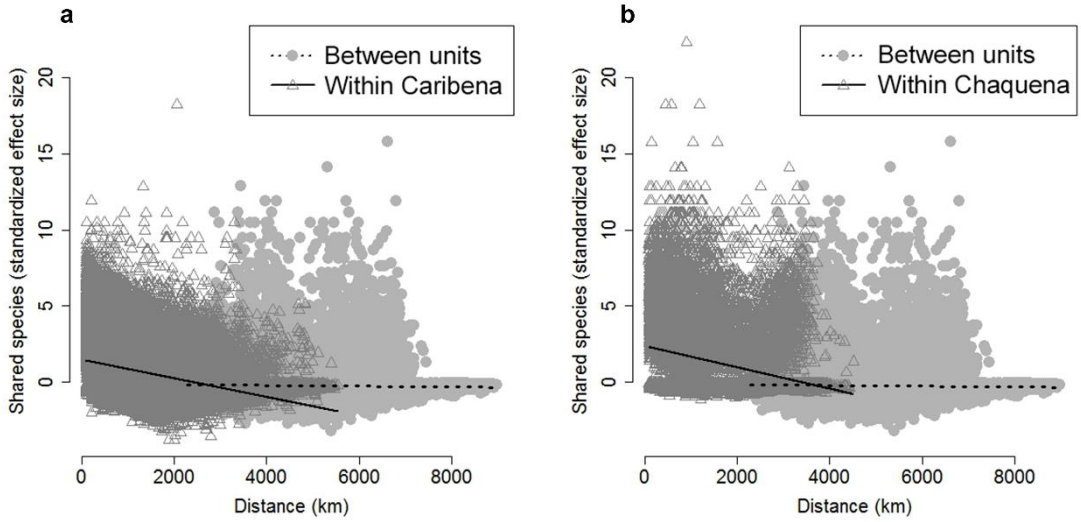


Figure 1.7. Comparisons of beta-diversity within and between two subregions of the Neotropical region: Caribena and Chaquena.
At short geographic distances beta-diversity between subregions Caribena and Chaquena was higher than beta-diversity within subregion Caribena (a) or Chaquena (b), but the pattern disappeared at larger geographic distances. Lines are matrix regression lines and each point represents a pair of 1-degree cells.

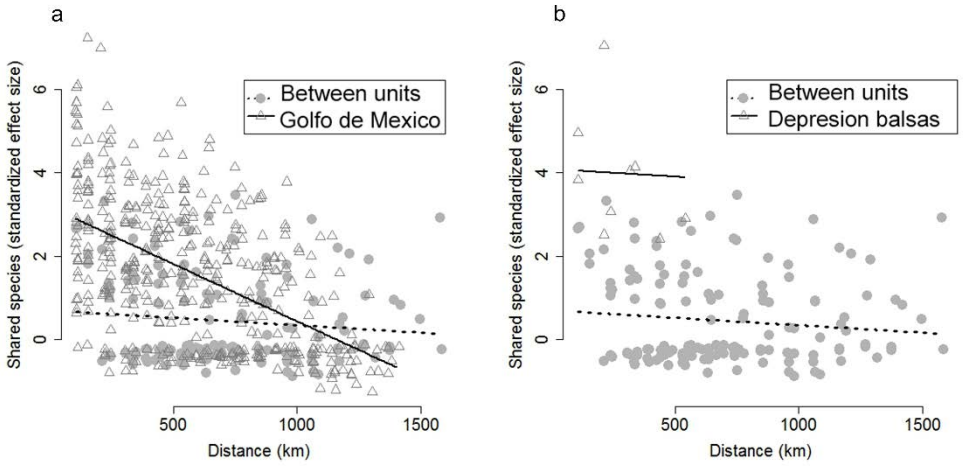


Figure 1.8. Comparison of beta-diversity within and between two provinces of subregion Caribena: Depresion de Balsas and Golfo de Mexico.
a) At short geographic distances beta-diversity between Depresion de Balsas and Golfo de Mexico was higher than beta-diversity within Golfo de Mexico, but the pattern disappeared at larger geographic distances, thus failing to support the prediction of biotic homogeneity: that variation in species composition between sites within a biogeographic unit should be lower than that between sites located in different biogeographic units. b) For all distances considered, beta-diversity between Depresion de Balsas and Golfo de Mexico was higher than beta-diversity within Depresion de Balsas. Thus, in contrast with the results shown in panel (a), the results in panel (b) support the prediction of biotic homogeneity. Symbols represent pairs of 1-degree cells, and lines are matrix regression lines.

As for the Caribena subregion, in 62 out 253 pairwise comparisons beta-diversity was lower within the same province than beta-diversity between provinces only at short geographic distances (Appendix 1.2). For instance, beta-diversity between the Golfo de Mexico province and each of 11 other provinces was higher than beta-diversity within Golfo de Mexico at short geographic distances (e.g., Fig. 1.8). Likewise, beta-diversity between Oriente de America Central and each of 10 other provinces was higher than beta-diversity within Oriente de America Central at short geographic distances.

The same pattern was found in the Amazonica subregion for 64 out of 116 pairwise comparisons between provinces (e.g., Fig. 1.9). For example, in the case of the Imeri and Yungas provinces, nine out of 12 pairwise comparisons showed lower beta-diversity within units than between units only at relatively short geographic distances (e.g., Fig. 1.9-1.10). Finally, two out of eight pairwise comparisons of provinces within the subregion Chaquena, and one out of six within the subregion

Paraense, displayed lower beta-diversity within than between provinces only at short geographic distances (Appendix 1.2).

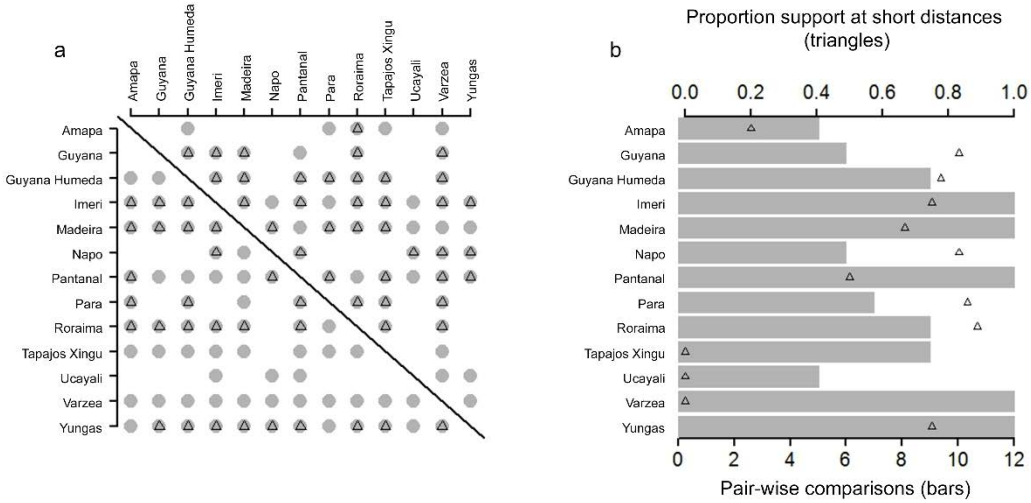


Figure 1.9. Summary of results for provinces in the subregion Amazonica in which beta diversity between provinces was higher than beta diversity within provinces only at short geographic distances.

a) Results for pairs of provinces in the subregion Amazonica. Each row shows the results for a single province paired to each of the provinces in the columns. Gray dots indicate that there was overlap between the geographic distances separating 1-degree cells located within the (row) province and the geographic distances separating 1-degree cells located in different provinces. Thus gray dots show the pairs of provinces for which the test of biotic homogeneity could be (and was) performed. Triangles indicate that beta-diversity within the (row) province is lower than beta-diversity between the pair of provinces at short geographic distances only, thus failing to supporting the prediction of biotic homogeneity. b) Summary of results for each province, showing the number of pairwise comparisons for which the test of biotic homogeneity could be (and was) performed (gray bars, lower axis) and the number of pairwise comparisons for which beta-diversity within the (row) province is lower than beta-diversity between the pair of provinces at short geographic distances only (triangles, upper axis).

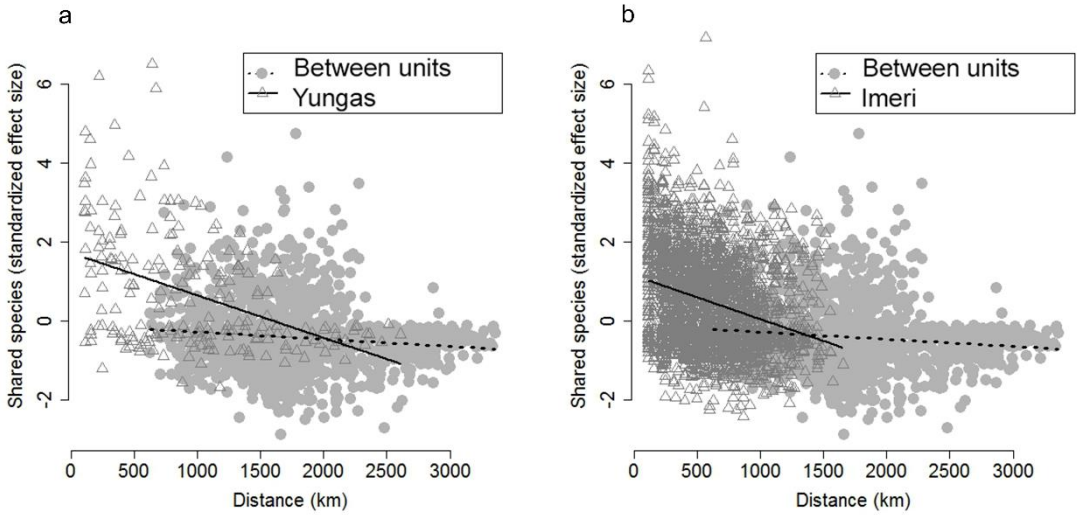


Figure 1.10. Comparison of beta-diversity within and between two provinces of subregion Amazonica: Yungas and Imeri.

a) At short geographic distances beta-diversity between Yungas and Imeri was higher than beta-diversity within Yungas, but the pattern disappeared at larger geographic distances. b) At short geographic distances beta-diversity between Yungas and Imeri was higher than beta-diversity within Imeri, but the pattern disappeared at larger geographic distances. Thus, the results shown in panels (a) and (b) fail to support the prediction of biotic homogeneity: that variation in species composition between sites within a biogeographic unit should be lower than that between sites located in different biogeographic units. Symbols represent pairs of 1-degree cells, and lines are matrix regression lines.

1.4 Discussion

Overall, our study did not empirically support the prediction of biotic homogeneity for the biogeographic units proposed by Morrone (2001). After accounting for the effect of distance and unequal botanical sampling, species beta-diversity between biogeographic units was higher than species beta-diversity within biogeographic units in only 113 out of 395 pairwise comparisons. The lack of biotic homogeneity within most of Morrone's units could be explained by the occurrence of various dispersal events which have driven distributional patterns in Sapotaceae (Armstrong *et al*., 2014; Richardson *et al*., 2014; Bartish *et al*., 2011). If the distribution of Sapotaceae has been largely affected by dispersal then a regionalization like that of Morrone (2001) based on cladistics and parsimony methods (Morrone, 2001, 2014), which assumes vicariance as the main mechanism driving patterns in the

assembly of communities (Brown and Lomolino, 1998; Crisp, 2006; Morrone, 2001, 2014; Lomolino *et al*., 2010; Nelson and Platnick, 1981), may not coincide with Sapotaceae's distributional patterns.

However, provinces like Chaco, Cuba, Varzea, Cauca, and Costa Pacifica Mexicana were biotically homogenous relative to several other provinces (Figs. 1.1, 1.4, 1.6, 1.11 and Appendix 1.1). Also, beta-diversity within units was lower than beta-diversity between units at relatively short geographic distances in 136 out 395 pairwise comparisons, including both provinces and subregions (e.g., Fig. 1.7 and 1.9).

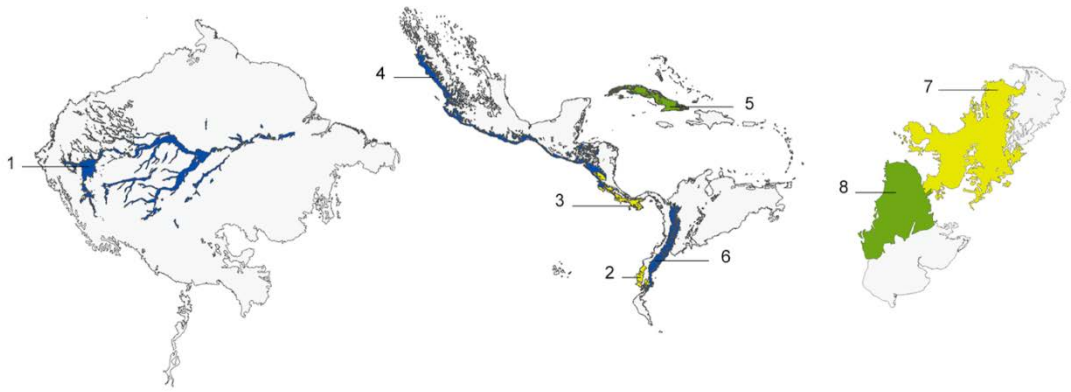


Figure 1.11. Ordinal representation of empirical support for provinces in the Caribena, Amazónica, Chaquena and Paranaense subregions Sensu Morrone (2001).
We did not find empirical support for the prediction of biotic homogeneity in terms of Sapotaceae species for the biogeographic units proposed by Morrone (2001). Nonetheless, biotic differences were found in >80% (in green) of the total pairwise comparisons performed in the provinces Cuba (5) and Chaco (8), in 70-80% (in blue) in Varzea (1), Cauca (6) and Costa Pacifica Mexicana (4); in 60-70% (in yellow) in Ecuador Arido (2), Occidente del Istmo de Panama (3) and Cerrado (7); and below 60% in the remaining units. Relatively high level of biotic homogeneity in Cuba, Cauca and Costa Pacifica Mexicana could be explained by isolation. In case of Varzea, patterns of aggregation of Sapotaceae species could have been driven by edaphic changes. Finally, Chaco, Ecuador Arido, Occidente del Istmo the Panama and Cerrado are units that include areas that are mostly dominated by grasslands and savannas, and are marginal for Sapotaceae's distributional range in which the highest number of species and individuals is found in tropical lowland rain forest.

1.4.1 Provincialism in neotropical Sapotaceae: dispersal and speciation

Biotic differences between several biogeographic units, where various pairwise comparisons showed lower beta-diversity within units than between them, even if homogeneity was not evident across all geographic distances or across all pairwise comparisons (e.g. Costa Pacifica Mexicana or Varzea), suggest that the biogeographic units proposed by Morrone (2001) capture at least some aspects of provincialism resulting from shared history, whereby dispersal is limited either by barriers, or species have locally diversified lacking enough time to colonize proximal areas.

The effects of mountain barriers could for example explain the relative high biotic homogeneity found in Costa Pacifica Mexicana (Figs. 1.1, 1.4, and 1.11). This province shares a border in the west with the Pacific Ocean and in the east with mountain ranges running along Central and North America e.g., The Sierra Madre mountain range in Mexico and Guatemala. These mountain barriers could affect community assembly in Costa Pacifica Mexicana by separating taxa on either side of their slopes. Communities in the lowland rain forest of these areas, as opposed to those in South America, are thought to have closer affinities to humid lowland montane forest in Central America than to other lowland rain forests in the Neotropics (Magallon *et al*., 2014). This could mean that families like Sapotaceae that occur in higher numbers below 1000 m of elevation and are species rich in lowland rain forests in other areas, are under-represented in units like the Costa Pacifica Mexicana.

Mountain barriers alone do not explain patterns of aggregation for Sapotaceae species in the Caribena subregion, however. We may have expected the Choco province (Fig. 1.1), also apparently isolated in this case by the Andean mountains, to show relatively high biotic homogeneity but this was not the case, suggesting that dispersal has prevented provincialism. Taxa in the Choco province would have been connected to the Central American flora by the closure of the Isthmus of Panama, and dispersal events between Choco and other provinces in inter-Andean valleys, or on the eastern side of the Andes in the Caribena subregion could have taken place across areas where the mountains are lower. In fact, the Choco and inter-Andean valleys of areas like Magdalena and Cauca are linked by lowland regions in the northern Caribbean region of Colombia and lowland passes adjacent to La Macarena and Norte de

Santander (Gentry, 1982). These areas of lowland are, however, occupied by or are adjacent to areas of dry forest or desert in the Caribbean and Magdalena valleys, respectively. The climates in these areas may have acted as barriers to overland dispersal of predominantly wet forest restricted Sapotaceae, but they may be fairly recently developed and not had time to have an effect on the distribution patterns in the family.

In the Cuba province on the other hand, biotic distinction (Fig. 1.1 and 1.11) could be explained by long periods of isolation, which may have allowed enough time for speciation since the island emerged in the middle Eocene (Graham, 2003; Iturralde-Vinent and MacPhee, 1999; Iturralde-Vinent, 1981). Cuba is the most plant species (ca. 6850 vascular plants) and endemic rich (ca. 3178 vascular plants) within the Greater Antilles. For instance, *Pouteria moaensis*, *P. cubensis, P. micrantha, P. aristata, Sideroxylon acunae, S. ekmanianum* and *Micropholis polita* are known Sapotaceae species endemic to this island (Martinez-Quezada, 2009; Figueredo, 2008; Pennington, 1990). Endemic flora in Cuba also includes taxa reported to have diversified at ca. 45 million years (Graham, 2003). This is particularly important specially if compared with other units like Jamaica or Haiti/Dominican Republic, which are of relatively similar size and also isolated, but are of younger origin.

Significant patterns of biotic homogeneity in terms of Sapotaceae species were also detected in the Amazonia subregion. Several pairwise comparisons showed lower beta-diversity within Varzea than between Varzea and other provinces (Fig. 1.1, 1.6 and 1.11), even if homogeneity was not evident across all geographic distances or across all pairwise comparisons. Relatively high biotic homogeneity of Varzea could be explained by abrupt edaphic changes between areas flooded by the Amazon River and *terra firme* away from the river that could act as adaptive barriers delimiting floral communities (Antonelli and Sanmartín, 2011). This would not be the case in other provinces within Amazonica where edaphic changes may be more localised (García-Villacorta *et al*., 2016) and not detectable using broad divisions like those of Morrone (2001).

Our analyses, also detected significant patterns and high values of biotic homogeneity in areas where Sapotaceae occurs but is not dominant, particularly in the

Chaco province. The high relative values of biotic homogeneity found in Chaco, could be partially a consequence of the occurrence of *Pouteria garderiana*. This species was reported by Morrone (2001) as a characteristic taxon for the Chaquena subregion, and in our data set, within this subregion, it was found more abundantly in the Chaco province.

In addition to biotic distinction of the Cauca, Costa Pacifica Mexicana, Cuba and Varzea and Chaco provinces, in several comparisons among subregions and in 62 out of 253 pairwise comparison among provinces beta-diversity was higher between units than within units at short geographic distances only (e.g., Figs. 1.7, 1.9, 1.10 and Appendix 1.2). This pattern suggests that the limits of some subregions and provinces *sensu* Morrone (2001) correspond to true biogeographic borders in terms of the distribution of Sapotaceae, but that there may be additional discontinuities within these biogeographic units that have not been identified.

An alternative scenario could attribute these discontinuities to the effects of distance decay (Soininen *et al*., 2007). However, in our study that seems unlikely considering that we controlled for this expected decrease in biotic similarity as geographic distance increases. In other words, in addition to controlling for differences in collection effort our analysis included geographic distance as an explanatory variable (D_{ij} in equation 1). Further work may explore in detail why a relatively high number of pairwise comparisons between biogeographic units show biotic homogeneity only at short geographic distances.

1.4.2 Methodological considerations

Lack of support for biotic homogeneity of units in Morrone (2001) regionalization could be primarily the result of dispersal events having an important effect in the distribution of Sapotaceae, but it is also likely influenced by our limited understanding about the world's biodiversity and its geographic characteristics i.e. the Wallacean and Linnaean shortfalls (see Introduction). In the present work, we attempted to control the effects of these variables by implementing a null model that accounts for differences in collection effort across 1-degree grid cells (sampling units). Specifically, our null model aimed to account for the concentration of collection effort in areas like northern Amazonia and northern Peru/southern Colombia, and relatively

low collection effort in areas possibly rich in Sapotaceae species, like the Choco biodiversity hotspot (Fig. 1.2). We believe few studies have accounted for geographic heterogeneity of collection effort in the past (but see Raup and Crick, 1979), but we also recognise that Sapotaceae species are likely to be discovered in the future and that depiction of distributional ranges of species already known to science is bound to change as botanists explore sparingly collected neotropical regions.

The recognition of patterns in the distribution of taxa is also strongly influenced by the choice of metrics to assess biotic similarity. Metrics for biotic similarity have been commonly developed combining variables to measure patterns of nestedness and species turnover. This approach has proven useful in popular indices like Jaccard and Sorensen. Nonetheless, nestedness and species turnover are the result of different processes that may not be clearly assessed if they are interpreted simultaneously (Mouillot *et al*., 2013). The aim of our analyses was to test if the limits of biogeographic units proposed in Morrone's biogeographic regionalization correspond to spatial changes in the occurrence of Sapotaceae species beyond what one may expect from distance decay of biotic similarity. These units were delimited in a hierarchical system of nested areas. Therefore, nested patterns were incorporated in our analysis implicitly in the metric of biotic similarity (i.e., standardized effect size for the number of shared species, SES_{ij} in equation 1) through the definition of "regional pools" for null models (see Methods). It is beyond the scope of this study to assess the degree to which biogeographic units are truly nested.

1.4.3 Sapotaceae as a study group

Biogeographic regionalizations, like that of Morrone (2001), have been often proposed using information on a few selected taxa (Lomolino, 1998), but ideally their predictions should be tested using data on additional groups of organisms (Magnusson, 2004, Whittaker *et al*., 2005). Although the biogeographic regionalization proposed by Morrone (2001) was partly based on information on the distribution of vascular plants, Sapotaceae do not seem to figure prominently among the taxa used to characterize different biogeographic units, as opposed to other plant families such as Burseraceae, Gunneraceae, Onagraceae, Passifloraceae and Podocarpaceae. Moreover, the Sapotaceae dataset used in this study (comprising 22,217 records

representing >95 of Sapotaecae diversity in the Neotropics) was recently compiled and curated to achieve relatively high spatial and taxonomic resolution, and thus was not available for the delineation of biogeographic units by Morrone (2001).

The test of biotic homogeneity presented here, based on data on the distribution of Sapotaceae species is largely independent from the original data used by Morrone (2001). However, it seems reasonable to use Sapotaceae species as exemplar taxa because they are important representatives of tropical lowland rain forest, which cover extensive areas within the Neotropics. Additionally, according to our results, Sapotaceae seem to be also an appropriate group for recognising patterns in habitats considered marginal for its distribution. For instance, patterns of biotic distinction were found in units dominated by grasslands and savannahs, like Chaco, Ecuador Arido, Occidente del Istmo de Panama and Cerrado.

It is certainly possible that future tests performed with data on other taxa will support the biogeographic regionalizations proposed by Morrone (2001), and that Sapotaceae prove to be an exception. Different taxa may have distinct evolutionary histories that determine distinct current distribution patterns (Proches, 2006). There is ample room for further progress in determining distributional patterns in the Neotropics, and future studies using alternative biogeographic regionalizations as hypotheses and additional taxa could provide more insights on the effects of dispersal in the distribution of taxa.

References

Anderson, M. J., Crist, T. O., Chase, J. M., Vellend, M., Inouye, B. D., Freestone, A. L., …Swenson, N. G. 2011. Navigating the multiple meanings of betadiversity: a roadmap for the practicing ecologist. Ecology Letters. 14: 19-28.

Antonelli, A. & Sanmartín, I. 2011. Why are there so many plant species in the Neotropics? Taxon. 60: 403-414.

Armstrong, K. E., Stone, G. N., Nicholls, J. A., Valderrama Escallón, E., Anderberg, A. A., Smedmark, J. & Richardson, J. E. 2014. Patterns of diversification amongst tropical regions compared: A case study in Sapotaceae. Frontiers in Genetics. 5: 1-13.

Bartish, I., Richardson, J. & Swenson, U. 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). Journal of Biogeography. 38: 177-190.

Brown, J. H. & Lomolino, M. V. 1998. Biogeography. 704.

Burnham, R. J., Graham, A., & Burnham, J. 2014. The History of neotropical Vegetation: New Developments and Status. 86: 546-589.

Burnham, R.J. & Johnson, K.R. 2004. South American paleobotany and the origins of neotropical rainforests. Philosophical Transactions of the Royal Society of London, series B. 359: 1595-1610.

Candolle, A. P. de. 1820. Géographie botanique. In: Dictionnaire des Sciences Naturelles. 18: 359-422.

Chase, J.M. & Mayers, J.A. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Philosophical Transactions of the Royal Society. 366: 2351-2363.

Cracraft, J. 1994. Species diversity, biogeography and the evolution of biotas. American Zoologist. 34: 33-47.

Crawley, M. J. 2002. Statistical Computing: An Introduction to Data Analysis using S-Plus. Chichester, West Sussex, England: Wiley. 772.

Crisp, M. 2006. Biome assembly: What we know and what we need to know. Journal of Biogeography. 33: 1332-1333.

Escalante, T. 2009. Un ensayo sobre regionalización biogeográfica. Revista Mexicana de Biodiversidad. 80: 551-560.

Essl, F., W. Rabitsch, S. Dullinger, D. Moser & N. Milasowszky. 2013. How well do we know species richness in a well-known continent? Temporal patterns of endemic and widespread species descriptions in the European fauna. Global Ecology and Biogeography. 22: 29-39.

Figueredo, L. M. 2008. Objetos de conservación de la flora y la vegetación de los cerros calizos costeros de la reserva de la biosfera Baconao, Santiago de cuba. Foresta Veracruzana. 10: 9-16.

Fortin, M. J. & Dale, M. R. T. 2005. Spatial analysis, a guide for ecologists. Cambridge: Cambridge University Press.

García-Villacorta, R., Dexter, K.G. & Pennington, R.T. 2016. Amazonian white-sand forests show strong floristic links with surrounding oligotrophic habitats and the Guiana shield. Biotropica. 48: 47-57.

Gentry, A. H. 1982. Phytogeographic patterns as evidence for a Chocó refuge. In: Prance, G.T. (ed.) Biological Diversification in the tropics. Columbia University Press, New York. 112-136.

GBIF.org (12th April 2015) GBIF Occurrence Download.

Govaerts, R. & Frodin, D. 2001. World Checklist and Bibliography of Sapotaceae, Royal Botanic Gardens, Kew. 372.

Graham, A. H. 2003. The Concepts of Deep-Time Floras and Paleobotanical Hot-Spots. Systematic Botany. 28: 461-464.

Holt, B. G., Lessard, J., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., … Jønsson, K.A. 2013. An Update of Wallace's Zoogeographic Regions of the World. Science, 339: 74-78.

Iturralde-Vinent M. A. 1981. Nuevo modelo interpretativo de la evolución geológica de Cuba. Ciencias de la Tierra y del Espacio. 3: 51-89.

Iturralde-Vinent, M. & MacPhee, R. D. 1999. Paleogeography of the Caribbean Region: Implications for Cenozoic Biogeography. Bulletin of the American Museum of Natural History. 238: 1-95.

Kreft, H. & Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. Journal of Biogeography. 37: 2029-2053.

Ladle, R. J. & Whitaker, R. J. 2011. Conservation Biogeography. Wiley-Blackwell. 320.

Legendre, P., & Legendre, L. 1988. Numerical Ecology (Developments in Environmental Modelling, 20). Second Edition Vol. 24. Quebec: Elsevier.

Loiselle, B. A., Jørgensen, P. M., Consiglio, T., Jiménez, I., Blake, J. G., Lohmann, L. G. & Montiel, O. M. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? Journal of Biogeography. 35: 105-116.

Lomolino, M. V., Riddle & Brown, J. H., 2006. Biogeography. Third edition. Sinauer Associates, Sunderland, MA. 752.

Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. H. 2010. Biogeography. Fourth edition. Sunderland, MA: Sinauer Associates. 878.

Mackey, B. G., Berry, S. L. & Brown, T. 2008. Reconciling approaches to biogeographical regionalization: a systematic and generic framework examined with a case study of the Australian continent. Journal of Biogeography. 35: 213-229.

McLaughlin, S. P. 1992. Are floristic areas hierarchically arranged? Journal of Biogeography. 19: 21-32.

Magallon, S., Pennington, R. T., Villasenor, J. L., Rodriguez, P., Honorio, E., Foster, R. & Wieringa, J. J. 2014. The influence of regional history and ecological affinity in the angiosperm composition of Mexican lowland tropical forests in Stevens, W., Montiel, O. M. & Raven, P. (ed.) Paleobotany and Biogeography. Missouri Botanical Garden Press. Missouri. 287-319.

Magnusson, W. E. 2004. Ecoregion as a pragmatic tool. Conservation Biology. 18: 4-6.

Martinez-Quezada, E. 2009. Phytogeographic relationships of spermatophytes on the reaindorest of the eastern region of Cuba. Gayana Botanica. 66: 177-205.

Morrone, J. 2014. Parsimony analysis of endemicity (PAE) revisited. Journal of Biogeography. 41: 842-854.

Morrone, J. 2001. Biogeografia de America Latina y el Caribe. First ed. Zaragoza: CYTED, ORCYT-UNESCO, Sociedad Entomológica Aragonesa (SEA).

Mouillot, D., J. D. Bortoli, F., Leprieur, V. Parravicini, M. Kullbicki & Bellwood, D. R. 2013. The challenge of delineating biogeographical regions: nestedness matters for indo-Pacific coral reef fishes. Journal of Biogeography. 40: 2228-2237.

Nelson, G. J. & Platnick, N. I. 1981. Systematics and Biogeography: Cladistics and Vicariance Columbia University Press, New York. 567.

Oliveiro, J., Márquez, A. L. & Real, R. 2013. Integrating fuzzy logic and statistics to improve the reliable delimitation of biogeographic regions and transition zones. Systematic Biology. 62: 1-21.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., … Morrison, J.C. 2001. Terrestrial ecoregions of the world: a new map of life on earth. Bioscience. 51: 933-938.

Olson, D. M., & Dinerstein, E. 2002. The Global 200: priority ecoregions for global conservation. Annals of the Missouri Botanical Garden. 89: 199-224.

Padme.rbge.org.uk. Royal Botanic Garden Edinburgh. 10 Apr 2014 <www.padme.rbge.org.uk>

Pennington, T. D. P. 2007. Sapotaceae In: G. Harling & C. Persson (ed.), Flora of Ecuador. vol. 152. Department of Plant and Environmental Sciences, Göteborg University, Sweden.

Pennington. T. D. P. 1990. Flora Neotropica, Vol. 52, Sapotaceae. New York Botanical Garden Press. 1-770.

Pennington. T. D. P. 1991. The genera of Sapotaceae. Royal Botanic Gardens Kew, New York Botanical Garden. 307.

Proches, S. 2006. Latitudinal and longitudinal barriers in global biogeography. Biology Letters. 2: 69-72.

Raup, D. M., & Crick, R. E. 1979. Paleontological Society Measurement of Faunal Similarity in Paleontology. Journal of Paleontology. 53: 1213-1227.

Richardson, J. E., Bakar, A. M., Tosh, J., Armstrong, K. E., Smedmark, J., Anderberg, A., Slik, F., Wilkie, P. 2014. The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae). Botanical Journal of the Linnean Society. 174: 13-140.

Sclater, P. L. 1858. On the general geographic distribution of the members of the class Aves. Journal of the Linnean Society. Zoology. 2: 130-145.

Sheth, S., Consiglio, T., Lohmann, L. G. & Jiménez, I. 2012. Understanding bias in geographic range size estimates. Global Ecology and Biogeography. 21: 732-742.

Soininen, J., McDonald, R. & Hillebrand, H. 2007. The distance decay of similarity in ecological communities. Ecography. 30: 3-12.

Stoddart, D. R. 1992. Biogeography of the Tropical Pacific. Pacific Science. 46: 276-293.

Stuart, Y. E., Losos, J. B. & Algar, A. C. 2012. The island-mainland species turnover relationship. Proceedings of the Royal Society of London B. 279: 4071-4077.

Swenson, U. & Anderberg, A. 2005. Phylogeny, character evolution, and classification of Sapotaceae (Ericales). Cladistics. 21: 101-130.

Swenson, U., Richardson, J. & Bartish, I. 2008. Multi-gene phylogeny of the pantropical subfamily Chrysophylloideae (Sapotaceae): Evidence of generic polyphyly and extensive morphological homoplasy. Cladistics. 24: 1006-1031.

Szumik, C. A. & Goloboff, P. 2004. Areas of endemism: an improved optimality criterion. Systematic Biology. 53: 968-977.

ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomao, R. P., Guevara, J. E., … Silman, M. R. 2013. Hyperdominance in the Amazonian Tree Flora. Science. 342: 6156.

Tropicos.org. Missouri Botanical Garden. 16 Mar 2015 <http://www.tropicos.org>

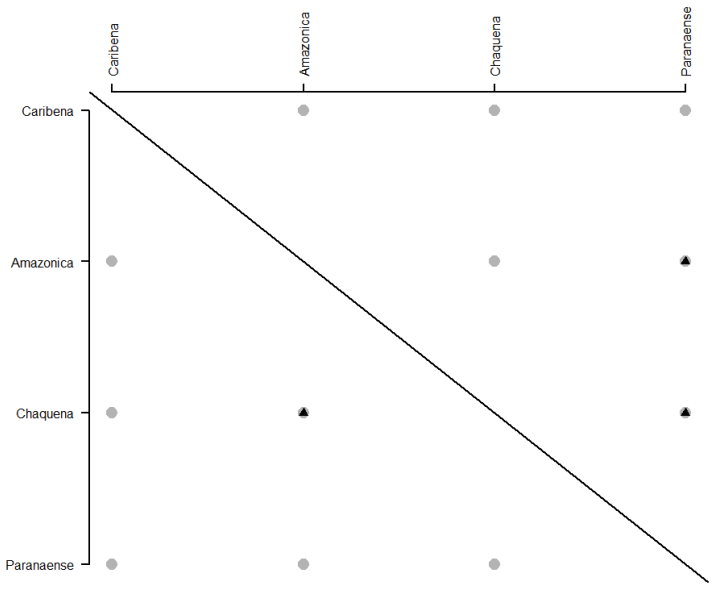
Vilhena, D. A, & Antonelli, A. 2015. A network approach for identifying and delimiting biogeographical regions. Nature Communications. 6: 1-9.

Wallace, A. R., 1876. The Geographical distribution of animals. London: MacMillan.

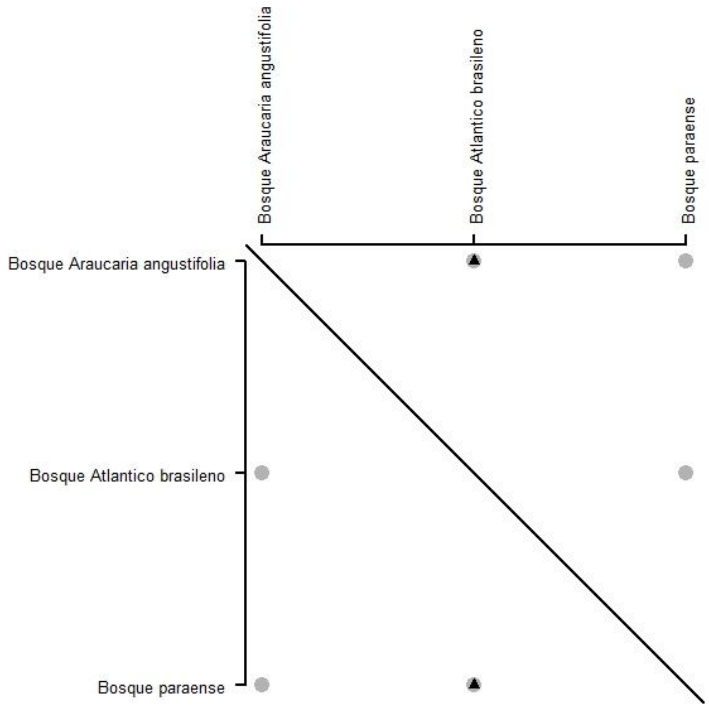
Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E. M. & Willis, K. J. 2005. Conservation biogeography: Assessment and prospect. Diversity and Distributions. 11: 3-23.

Appendices

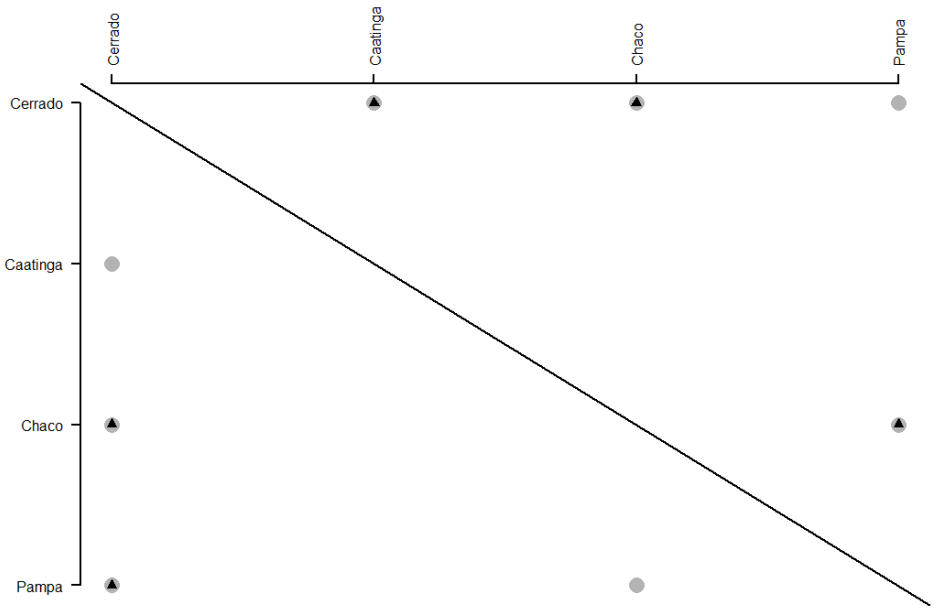
Appendix 1.1



a



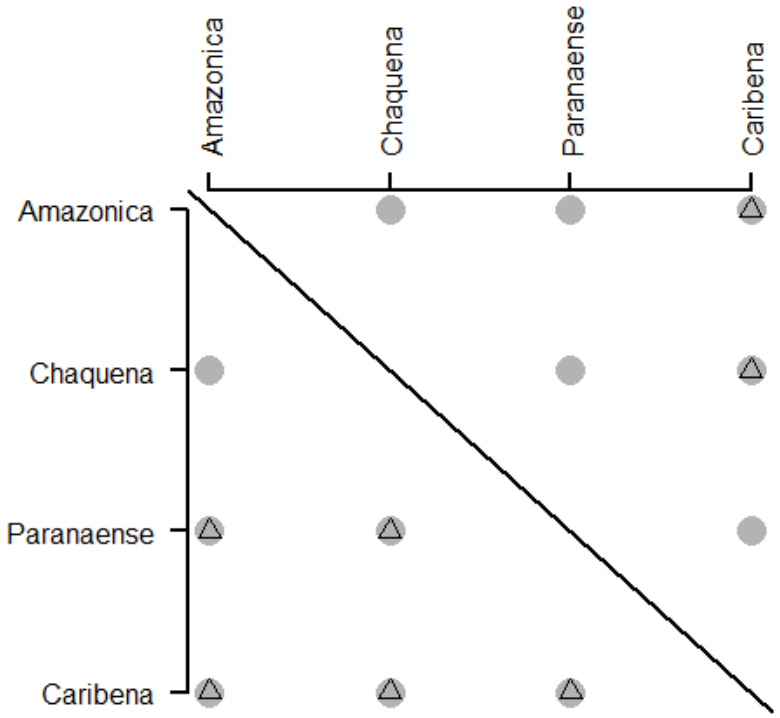
b



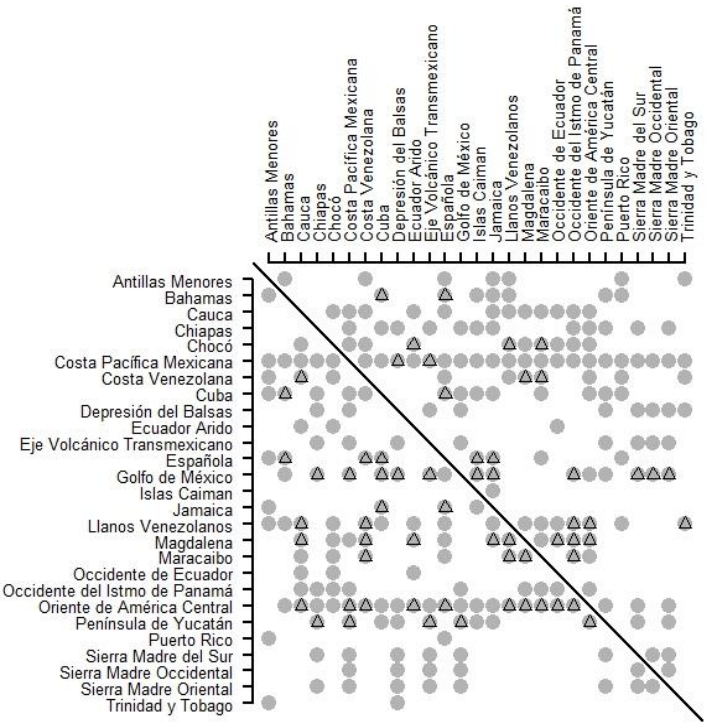
c

Appendix 1.1. Summary of results from the test of biotic homogeneity for subregions in the Neotropical region.
(a), provinces in the subregion Paranaense (b), and Chaquena (c). Each row shows the results of tests of biotic homogeneity of a particular biogeographic unit (i.e., the unit after which the row is labelled) relative to each of the biogeographic units in the columns. Gray dots indicate that there was overlap between the geographic distances separating 1-degree cells located within the (row) biogeographic unit and the geographic distances separating 1-degree cells located in different biogeographic units. Thus gray dots show the pairs of biogeographic units for which the test of biotic homogeneity could be (and was) performed. Triangles indicate that beta-diversity within the (row) biogeographic unit was lower than beta-diversity between the pair of biogeographic units, thus empirically supporting the prediction of biotic homogeneity.

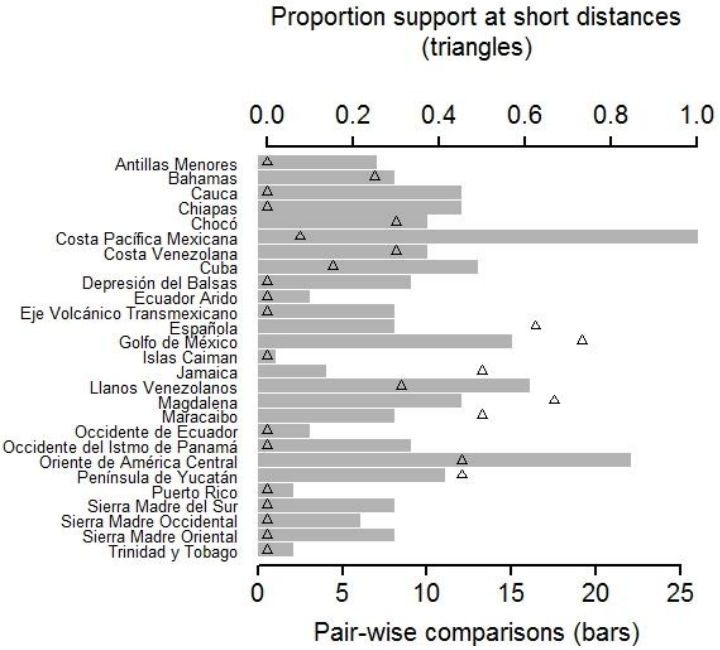
Appendix 1.2



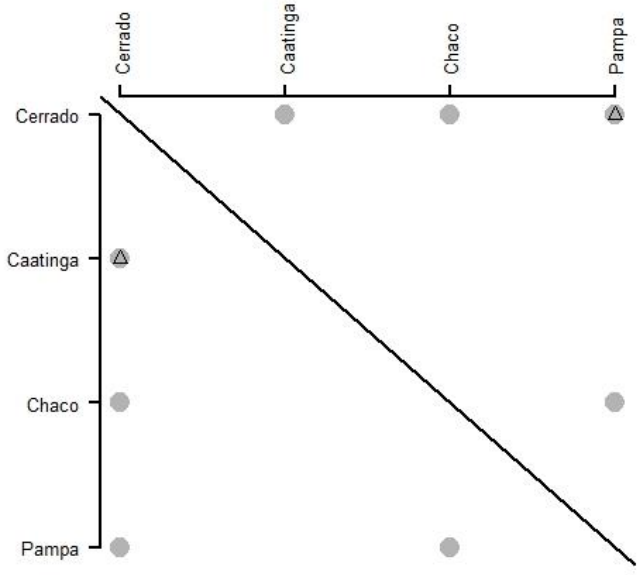
a



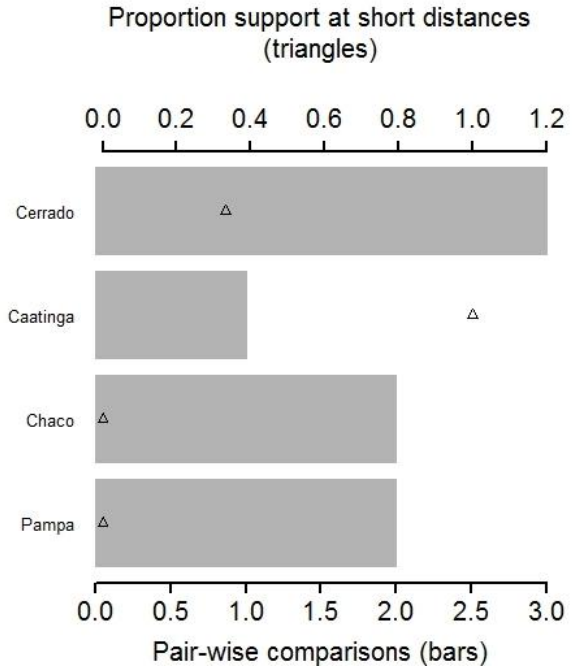
b


Proportion support at short distances
(triangles)
0.0
0.2
0.4
0.6
0.8
1.0
Antillas Menores
Bahamas
Cauca
Chiapas
Chocó
Costa Pacífica Mexicana
Costa Venezolana
Cuba
Depresión del Balsas
Ecuador Árido
Eje Volcánico Transmexicano
Española
Golfo de México
Islas Caiman
Jamaica
Llanos Venezolanos
Magdalena
Maracaibo
Occidente de Ecuador
Occidente del Istmo de Panamá
Oriente de América Central
Península de Yucatán
Puerto Rico
Sierra Madre del Sur
Sierra Madre Occidental
Sierra Madre Oriental
Trinidad y Tobago
0
5
10
15
20
25
Pair-wise comparisons (bars)

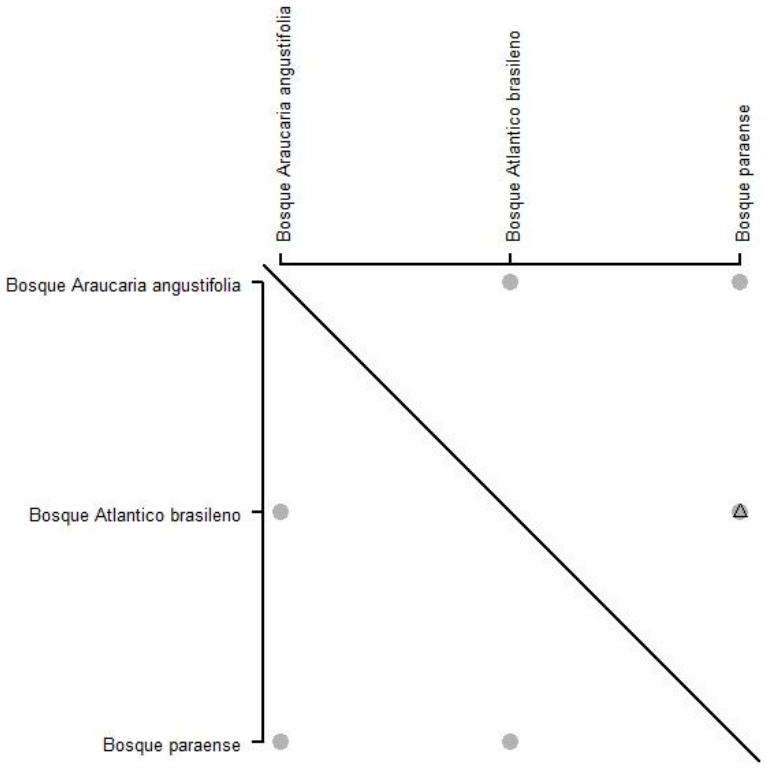
c


Cerrado
Caatinga
Chaco
Pampa
Cerrado
Caatinga
Chaco
Pampa

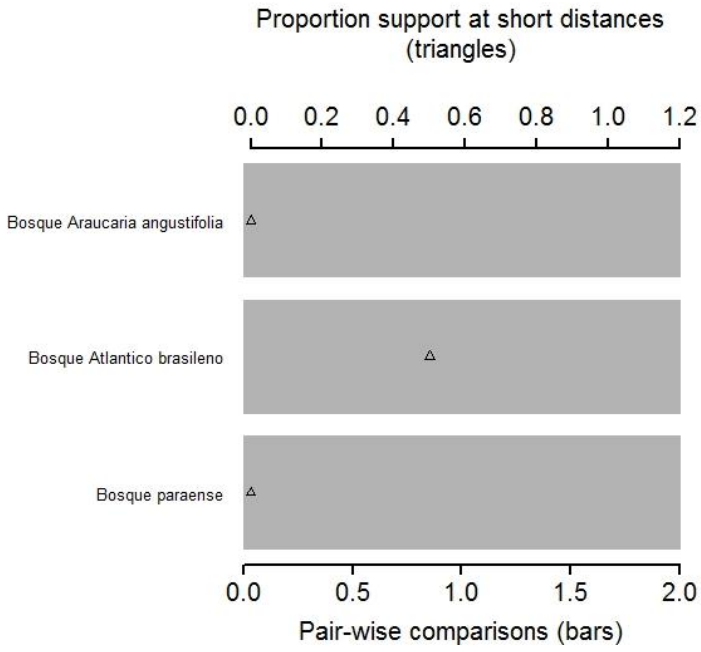
d


Proportion support at short distances
(triangles)
0.0 0.2 0.4 0.6 0.8 1.0 1.2
Cerrado
Caatinga
Chaco
Pampa
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Pair-wise comparisons (bars)

e


Bosque Araucaria angustifolia
Bosque Atlantico brasileno
Bosque paraense
Bosque Araucaria angustifolia
Bosque Atlantico brasileno
Bosque paraense

f



g

Appendix 1.2. Summary of results for the Neotropical region.
Summary of results for subregions in the Neotropical region (a), and Provinces in the subregion Caribena (b and c), Chaquena (d and e) and Paranaense (f and g) in which beta diversity between units was higher than beta diversity within units only at short geographic distances. a, b, d and f) Results for pairs of units. Each row shows the results for a single unit paired to each of the units in the columns. Gray dots indicate that there was overlap between the geographic distances separating 1-degree cells located within the (row) unit and the geographic distances separating 1-degree cells located in different units. Thus gray dots show the pairs of units for which the test of biotic homogeneity could be (and was) performed. Triangles indicate that beta-diversity within the (row) unit is lower than beta-diversity between the pair of units at short geographic distances only, thus failing to supporting the prediction of biotic homogeneity. c, e and g) Summary of results for each unit, showing the number of pairwise comparisons for which the test of biotic homogeneity could be (and was) performed (gray bars, lower axis) and the number of pairwise comparisons for which beta-diversity within the (row) unit is lower than beta-diversity between the pair of units at short geographic distances only (triangles, upper axis).

Chapter 2. Effects of Geological and Climatic Changes on the Diversification of Lowland Neotropical Rain Forest Trees: A Case Study in Sapotaceae

2.1 Introduction

The neotropical lowland rain forest is a species-rich biome thought to have originated at the start of the Cenozoic (Burnham *et al*., 1999). In northern South America it is represented by several areas of forest, for example the Chocó, Amazon, Catatumbo, Macarena and some fragments in the inter-Andean valleys (Fig. 2.1). These areas have been documented to host high numbers of species, and they are considered of great conservation and economic importance (Bernal *et al*., 2016), yet the origin and distributional patterns of this diversity are still unclear. Currently, evidence suggests that distributional patterns of plant species in neotropical lowland rain forest have been affected by both vicariance and long distance dispersal events (Burnham *et al.,* 1999; Pennington *et al.,* 2004) and influenced by the uplift of the Andean mountains, the closure of the Panama Isthmus and climatic changes during the Pleistocene (Burnham *et al.,* 1999; Pennington *et al.,* 2004; Renner *et al.,* 2001; Renner, 2004; Richardson *et al.,* 2004; Antonelli *et al.,* 2009; Duangjai *et al.,* 2009; Winterton *et al.,* 2014). The main purpose of this chapter is to infer the impact of these processes on the biogeography of Sapotaceae in northern South America, using a molecular phylogeny and the first data set for this family that samples numerous accessions from lowland rain forests in Colombia.

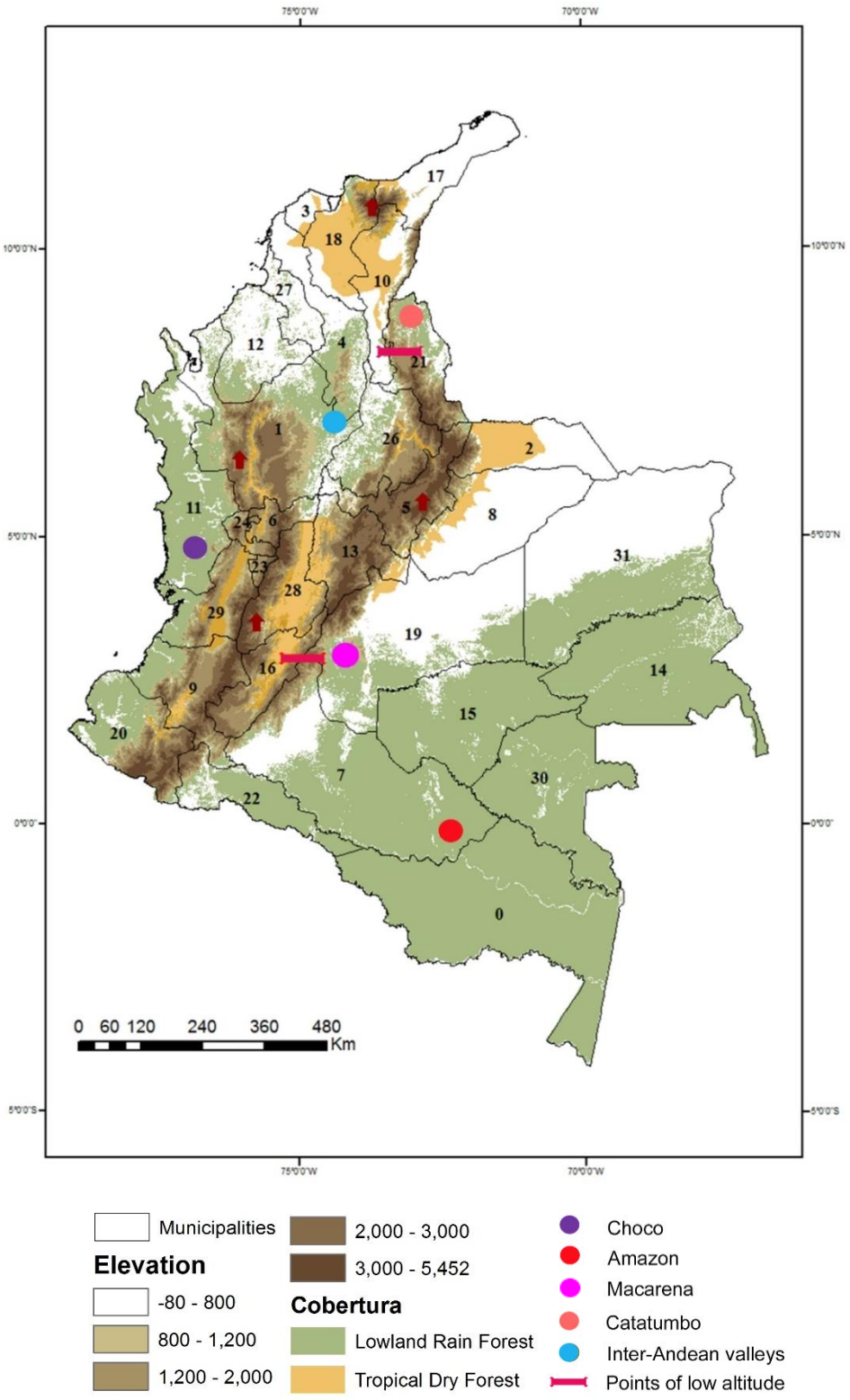

75°0'0"W
70°0'0"W
10°0'0"N
5°0'0"N
0°0'0"
5°0'0"S
17
3
18
10
27
12
4
21
1
26
2
11
24
6
5
8
13
31
23
28
29
19
16
14
9
15
20
7
30
22
0
0 60 120 240 360 480
Km
Municipalities
Elevation
-80 - 800
800 - 1,200
1,200 - 2,000
2,000 - 3,000
3,000 - 5,452
Cobertura
Lowland Rain Forest
Tropical Dry Forest
Choco
Amazon
Macarena
Catatumbo
Inter-Andean valleys
Points of low altitude

Figure 2.1. Lowland rain forests of Colombia. Coloured dots represent the fragments of lowland rain forest that can be found in Colombia.
These areas are characterised by evergreen vegetation occurring under ca. 800 meters above sea level, which is dominated by angiosperm tree species adapted to low fluctuations in temperature, high precipitation (more than 2000 mm per year), with high beta and alpha diversity and an abundance of lianas and epiphytes (Olson et al., 2001; Burnham and Johnson, 2004; IDEAM et al., 2007; Reynel et al., 2013). Red arrows locate major mountain ranges, from north to south: Sierra Nevada de Santa Marta, Western, Eastern and Central Cordilleras. Black solid lines represent the admistrative division of Colombia into municipalities 0: Amazonas, 1: Antioquia, 2: Arauca, 3: Atlantico, 4: Bolivar, 5: Boyaca, 6: Caldas, 7: Caqueta, 8: Casanare, 9: Cauca, 10: Cesar, 11: Choco, 12: Cordoba, 13: Cundinamarca, 14: Guainia, 15: Guaviare, 16: Huila, 17: La Guajira, 18: Magdalena, 19: Meta, 20: Nariño, 21: Norte de Santander, 22: Putumayo, 23: Quindio, 24: Risaralda, 26: Santander, 27: Sucre, 28: Tolima, 29: Valle del Cauca, 30: Vaupes, 31: Vichada. Cover land units of Rain forest are depicted as proposed by IDEAM et al. (2007), and dry forest areas as adapted from the WWF Ecorregions (Olson et al., 2001).

2.1.1 Geological History of northern South America

The Andean orogeny took place from north to south and east to west in well differentiated phases (Gregory-Wodzicki, 2000; Mora *et al.,* 2010), but generally major tectonic movements started during the late Oligocene after the break-up of the Farallones plate, and the subsequent increment in convergence rates at the margins of the newly created Cocos and Nazca plates (Hoorn *et al.,* 1995; Wortel, 1984). The Andes divide into the Eastern, Central and Western Cordilleras in Colombia (Fig. 2.1). The Eastern Cordillera was formed as a continuous range in the mid Miocene (Winship, 1990; Hoorn *et al.,* 1995; Albert *et al.,* 2006). It reaches the northern coast of Colombia, branching into the Serrania del Perija and the Santander massif to the west, and into the Merida Andes to the east (Graham, 2011; Hoorn *et al.,* 1995). The section corresponding to the Santander massif is reported to have uplifted during the early and mid-Miocene (Hoorn *et al.,* 1995). The Merida Andes, on the other hand, have a younger origin and are reported to have risen above 3000 m at ca. 5-2 mya (Kroonenberg *et al.,* 1990). Currently the Merida Andes reach approximately 5,000 m in elevation enclosing the Maracaibo Basin (ca. 6,000 m in depth) (Kroonenberg *et al.,* 1990).

The Central Cordillera in Colombia reached its current altitude during the Miocene at ca. 10-4 mya (Kroonenberg *et al.,* 1990), creating altitudinal belts that at their lower points formed the Magdalena and the Cauca valley. This orogeny was partially contemporary to the accretion of the Western Cordillera in Colombia, a process that initiated during the Oligocene and early Miocene, and caused the creation

of high altitude habitats, above ca. 4,000 m, during the late Miocene (Graham, 2011; Kroonenberg *et al.,* 1990).

The uplift of the Andean mountains in the Miocene, particularly the Eastern Cordillera in northern South America, created a considerable physical barrier, and a dominant paradigm has considered it as having a significant impact on biogeographic patterns (Kroonenberg *et al.,* 1990; Hoorn *et al.,* 1995; Morrone, 2001; Cortés-Ortiz *et al.,* 2003; Gonzalez *et al.,* 2003; Albert *et al.,* 2006; Pirie *et al.,* 2006; Brumfield & Edwards, 2007; Hoorn *et al.,* 2010). The Central and Western Cordilleras in northern South America do not extend to the Caribbean coast and thus may not represent a barrier to dispersal of lowland restricted organisms. However, the Eastern Cordillera extends to coastal areas and therefore forms a near continuous barrier.

During the Miocene epoch the tectonic activity mentioned above, together with shifts in the Caribbean plate, resulted not only in the uplift of the Andes, but also the formation of the Panama land bridge (Hoorn *et al.,* 1995; Farris *et al.,* 2011). Both events dramatically altered the paleogeography, climate and possibly the distribution patterns of taxa in the Americas. They created novel habitats, altered fluvial systems, opened migration routes and caused habitat fragmentation (Hoorn, 2010, Hoorn *et al.,* 1995), for example in the Amazon and the Orinoco rivers. Prior to the Andean uplift, the younger Amazon River was only a fluvio-lacustrine system, initially not connected to the Atlantic Ocean. Instead, the mouth of the proto-Amazon was linked to what at the time was the Orinoco River. It was not until the late Miocene when continuous tectonic movements finally redirected the Amazon to the Atlantic Ocean and altered the course of the Orinoco River. Simultaneously, habitat fragmentation occurred due to the division of the previously continuous block formed by what at present are the lowland rain forests of the Amazon, the Magdalena Valley, and Choco. The isolation of these areas could have resulted in changes in distribution of lowland wet forest restricted lineages that may then have led to allopatric speciation, if these lineages were not capable of adapting to those changes.

Changes in the paleogeography and in the distribution of taxa in northern South America may also have occurred during the Pleistocene as a result of climatic oscillations during that epoch. Those changes may have resulted in the restriction of

wet forest to refuges and could have promoted allopatric speciation in taxa with low tolerances to fluctuations in temperature and precipitation (Haffer, 1969; Prance, 1973; Richardson, 2001 but see Colinvaux *et al.,* 2001; Pennington *et al.,* 2004 and Whinnett *et al.,* 2005).

2.1.2 Study group

Sapotaceae is a family of trees mainly restricted to lowland tropical rain forest. Currently, it has a pantropical distribution, but it is thought to have first diversified in Asia between ca. 67.1-105 mya (Richardson *et al.,* 2014). Three subfamilies are recognised within Sapotaceae: Chrysophylloideae, Sapotoideae and Sarcospermatoideae (Pennington, 1991), with Chrysophylloideae being the most abundant in the Neotropics. Bartish *et al.,* (2011) reported that the crown node of Chrysophylloideae was ca. 77.8 [72.9-82.7] mya and that the group originated in Africa. From Africa, the sub-family migrated to the Americas by a single event, suggested to be long distance trans-oceanic dispersal in the late Cretaceous (Bartish *et al.,* 2011). Chrysophylloideae, which are distributed almost exclusively in lowland rain forest, make a perfect model for studying the effects of various geological and climatic processes of the Tertiary and Quaternary epochs. Biogeographic studies have focused mainly in the Paleotropics, and patterns and processes in northern South America have been difficult to infer largely because of the lack of botanical collection in areas like the lowland rain forest of Colombia where Sapotaceae species are highly diverse.

The main aim of this chapter is to explore the diversification history of Sapotaceae, focussing on Chrysophylloideae, an important ecological component of the lowland rain forests in northern South America, and to relate this history to the Andean uplift, the closure of the Isthmus of Panama, and historical climatic changes (Fig. 2.2). To achieve this, phylogenetic sampling of Sapotaceae species on both sides of the Andes in Colombia and in the inter-Andean valleys was increased and used in a biogeographic reconstruction of the subfamily Chrysophylloideae using the internal transcribed spacer (ITS) region of nuclear ribosomal DNA. Given the restriction of Sapotaceae to lowland wet forest areas (below 1000 m a.s.l.) we might expect splits in the phylogeny to occur when the mountains reached a particular altitude (Fig. 2.2) or

when particular belts of dry vegetation were formed (possibly as a result of rain shadow effects due to Andean uplift). If Sapotaceae could not disperse across water we would also expect to see evidence of migration across the Isthmus of Panama only after its formation. Following the same logic of dispersal limitation, the refuge theory would predict that speciation would have occurred as a result of wet forest range restrictions during drier and cooler periods of the Pleistocene.

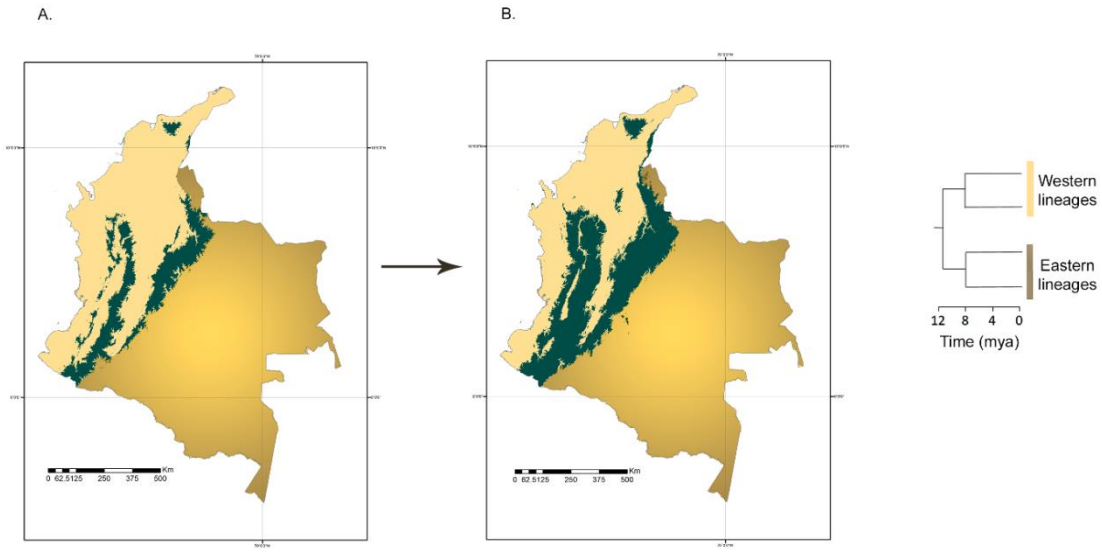


Figure 2.2. Hypothetical phylogenetic splits assuming speciation events due to the Andean uplift without subsequent dispersal.
Panel A. representing the Andean configuration and hypothetical species distribution (shades of yellow) before the early-middle Miocene, and Panel B. representing Andean configuration and hypothetical species distribution (shades of yellow) after early-middle Miocene when the Andes reached current elevations. If the evolution of Sapotaceae was affected by early and middle Miocene orogeny without subsequent dispersal, dichotomies in the phylogeny would occur at ca. 11.8 mya when the Eastern Cordillera was high enough to have prevented migration.

Barriers to dispersal could therefore be:

- The Central American Seaway prior to formation of the Isthmus of Panama. If Sapotaceae were not able to disperse across water we would predict migration across the Panama Isthmus only after its formation

- The Eastern Cordillera of the Andes. If mountains promoted allopatric speciation, we would expect splits to occur at the age at which those barriers were formed (e.g. the scenario illustrated in Fig. 2.2).
- The dry forests of Tatacoa and the southern Magdalena Valley (in the region of Departments 16 and 28 in Fig. 2.1) or Arauca (Department 2 in Fig. 2.1). If Sapotaceae were not able to migrate across dry forests we would expect splits in our phylogeny that coincided with the formation of those dry forests. We do not know the age at which these dry forests formed, hence the splits that occur based on our dated phylogeny may tell us something about the age of those dry forests.

2.2 Methods

A total of 146 new Chrysophylloideae accessions were collected in the Choco, Magdalena valley, Catatumbo, Amazon and Macarena lowland rain forests of Colombia, and were sequenced for the internal transcribed spacer of ribosomal DNA (ITS). The remaining sequences were obtained from previous phylogenetic studies on Chrysophylloideae (Sánchez-C. *et al.,* 2017; Gonzalez *et al.,* 2009; Swenson *et al.,* 2008) and GenBank. A list of taxa and voucher specimen information is shown in Appendix 2.1. ITS was chosen as it has been successfully used in biogeographic studies that used Sapotaceae as a model group (e.g. Bartish *et al*., 2011). ITS improved the resolution of those phylogenies, particularly among recently evolved taxa and closely related species.

2.2.1 DNA isolation, amplification and sequencing

Leaf tissue from fresh silica-gel-dried collections was ground into a powder. Total DNA was isolated with the DNeasy® Plant Mini Kit (QIAGEN) following the manufacturer's procedures. The nuclear ITS region was amplified using primer pairs ITS5 and ITS8. PCR reactions were carried out in 20µL volume reactions, by adding 2µl of dNTPs, 2 µl of 10x NH4 reaction buffer, 0.6 µl of $MgCl_2$, 1.5 µl of each primer (10µM), 4µl of CES, 0.3µl of Biotaq DNA polymerase buffer, 7.1 µl double distilled H2O (ddH2O) and 1µl of DNA. The thermal cycling profile was: 3 minutes at 94°C (denaturing) followed by 29 cycles of 1 minute at 94°C (denaturation), 1 minute at

55°C (annealing) and 1.5 minutes at 72°C (extension), and an additional termination step of 5 minutes at 72°C. The amplified fragments were electrophoresed on a 1% agarose gel to check for quantity. PCR products were then purified with ExoSAP-IT® (USB Corporation) which was added directly and incubated at 37°C for 15 minutes followed by 80°C for 15 minutes. PCR templates were used for the cycle-sequencing reactions using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) and then were run on an ABI PRISM 310 sequencer (Applied Biosystems, Inc.). One of the samples in this study (*Chrysophyllum colombianum*) had two bands on the PCR gel. These were excised and sequenced separately (*C. colombianum* a and b). We detected no other examples of multiple copies of ITS.

2.2.2 Sequence alignment and phylogenetic analysis

DNA sequences were assembled and edited using Geneious 10.1.2 (Kearse *et al.,* 2012), automatically aligned using ClustalW and edited manually with BioEdit (Hall, 1999). *Sarcosperma laurinum* was used as the outgroup in all analyses as it has been shown to be the sister to all other Sapotaceae (Anderberg and Swenson, 2003).

Maximum likelihood analyses were performed using RAxml (Stamatakis, 2014) on the CIPRES portal (www.phylo.org; Miller *et al.,* 2010). A general time reversible (GTR) with a gamma distribution and invariant sites was identified as the best-fit substitution model in Jmodeltest (Darriba *et al.,* 2012; Guindon and Gascuel, 2003). Analyses were set to run for a maximum of 24 hours with rapid bootstrapping and 1,000 iterations. The remaining settings were set to default.

Bayesian analyses were carried out with MrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) through the CIPRES gateway (www.phylo.org; Miller *et al.,* 2010). Four independent runs of four MCMC chains were set to run for 30,000,000 cycles, sampling the Markov chain every 3,000 cycles (10,000 samples) and saving branch lengths. After determining, using Tracer v3.2.6, that a burn-in of 25% sufficed to reach stationarity, the remaining samples were saved and used to construct a 50% majority-rule consensus tree. The average standard deviation was calculated every 30,000 generations, 1,000 calculations in total. Tree probabilities were calculated and credibility intervals were set as a fraction of the highest posterior density. Clade support was represented by posterior probability (pp.)

values, with pp values between 50 and 95% indicating weakly supported nodes and pp. values more than 95% indicating strong support (Swenson *et al.,* 2008). ESS values were checked in tracer and all parameters reached values >200.

2.2.3 Divergence time estimation

Divergence times were estimated using Bayesian dating analyses in BEAST v1.8.4 (Drummond and Rambaut, 2007). The XML (eXtensible Mark-up Language) input file was generated in BEAUti v1.8.4. The length of the chain was set to 40,000,000 cycles, logging parameters and printing to screen every 4000 cycles. An extra run of additional 40,000,000 cycles was set and combined to the first run to achieve ESS values higher than 200 for all parameters. A general time reversible (GTR) substitution model with a gamma distribution and invariant sites was used. Gamma categories were set to four and base frequencies were estimated. To allow for changes in rates along the branches an uncorrelated relaxed clock with a lognormal distribution was set. For the tree priors, we used the best tree from the maximum likelihood analyses as a starting tree, and a speciation birth-death process to account for background extinction. Priors for model parameters and statistics used a lognormal distribution for all fossil-based calibration points and a normal distribution for the secondary calibration point (Ho, 2007; Ho and Phillips, 2009).

Two fossil calibration points were used along with a secondary calibration point based on a dated phylogeny of a broader sample of asterids that utilized six well characterized fossils (Bremer *et al.,* 2004). *Psilatricolporites maculosus* (Chrysophylloideae) has a more or less continuous sequence at the late Paleocene-early Eocene transition in the Maracaibo Basin in western Venezuela (Rull, 1997, 2000). The record of *Psilatricolporites* from Venezuela is probably the oldest reliably dated representative of Chrysophylloideae from the New World. Therefore, an offset with an age estimate of 55 mya was used at the crown node that included all members in the New World (node B, Fig. 2.3). A mean of 0.001 was set so 95% of the intervals would include values between the estimated age and the latest boundary of the Late Paleocene. Pollen of *Psilastephanocolporites malacanthoides*, reported from the upper Eocene of Nigeria (Jan du Chêne *et al.,* 1978), was assigned to *Malacantha alnifolia*. The stem node of *M. alnifolia* was therefore constrained to an offset of 35 mya and a

mean of 0.001 (node C, Fig. 2.3), so that 95% of the intervals would include values between the estimated age and the latest boundary of the Late Eocene. Finally, the crown node of the phylogeny was dated using a secondary calibration point obtained from a dated phylogeny of asterids by Bremer *et al.,* (2004, 2009). This study estimated the age of the stem node of Sapotaceae to be 102 mya. Because we did not have any representatives of taxa that are sister to Sapotaceae in our analysis we could not recover the stem node. We therefore applied the age of the stem node of Sapotaceae to the crown node. We acknowledge this may bias our estimates towards the oldest possible age. The normal distribution was set with a mean of 102 and a standard deviation of three, so 95% of the intervals would include the age estimated by Bremer *et al.,* (2004, 2009) +/- five years (node A, Fig. 2.3).

2.2.4 Ancestral State Reconstruction

The RASP (Reconstruct Ancestral State in Phylogenies) software was used to perform BayArea and Bayesian Binary MCMC (BBM) (Yu *et al.,* 2015). Terminals were assigned to the following areas that were defined on the basis of isolation due to tectonics or by areas of dry climate: Amazon, Atlantic Forest, Catatumbo, Guyanas, Llanos, Macarena, West Indies, Central America, Choco, Cauca Valley, Magdalena Valley, Mexico, Africa, Argentina, Asia, and Australia (Fig. 2.3). The MCC tree obtained in the BEAST analyses was used as an input tree. A total of 5'000,000 cycles and ten chains sampling every 100 and discarding 5,000 samples were run. The maximum number of areas per node was set to four and the remaining settings were set to default.

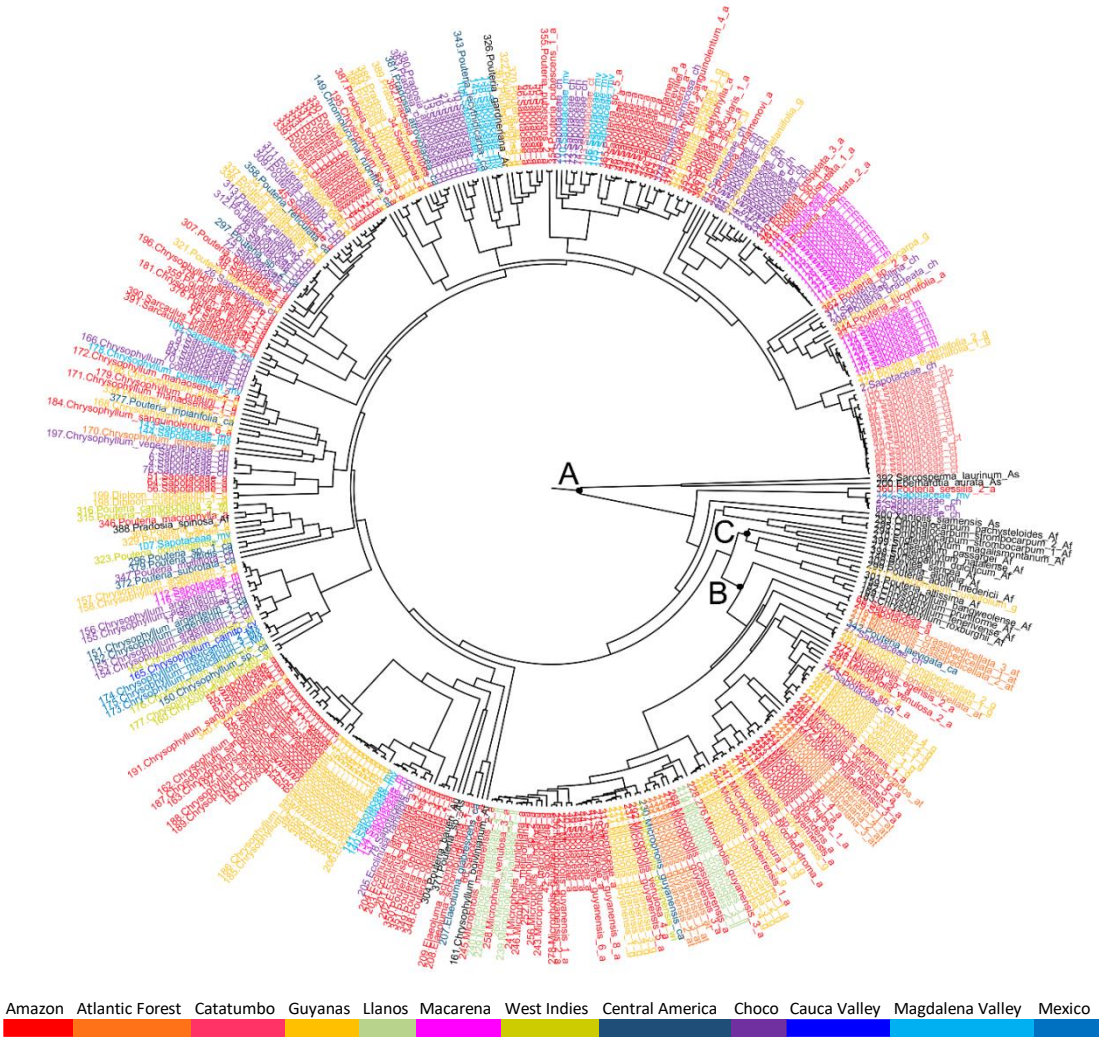


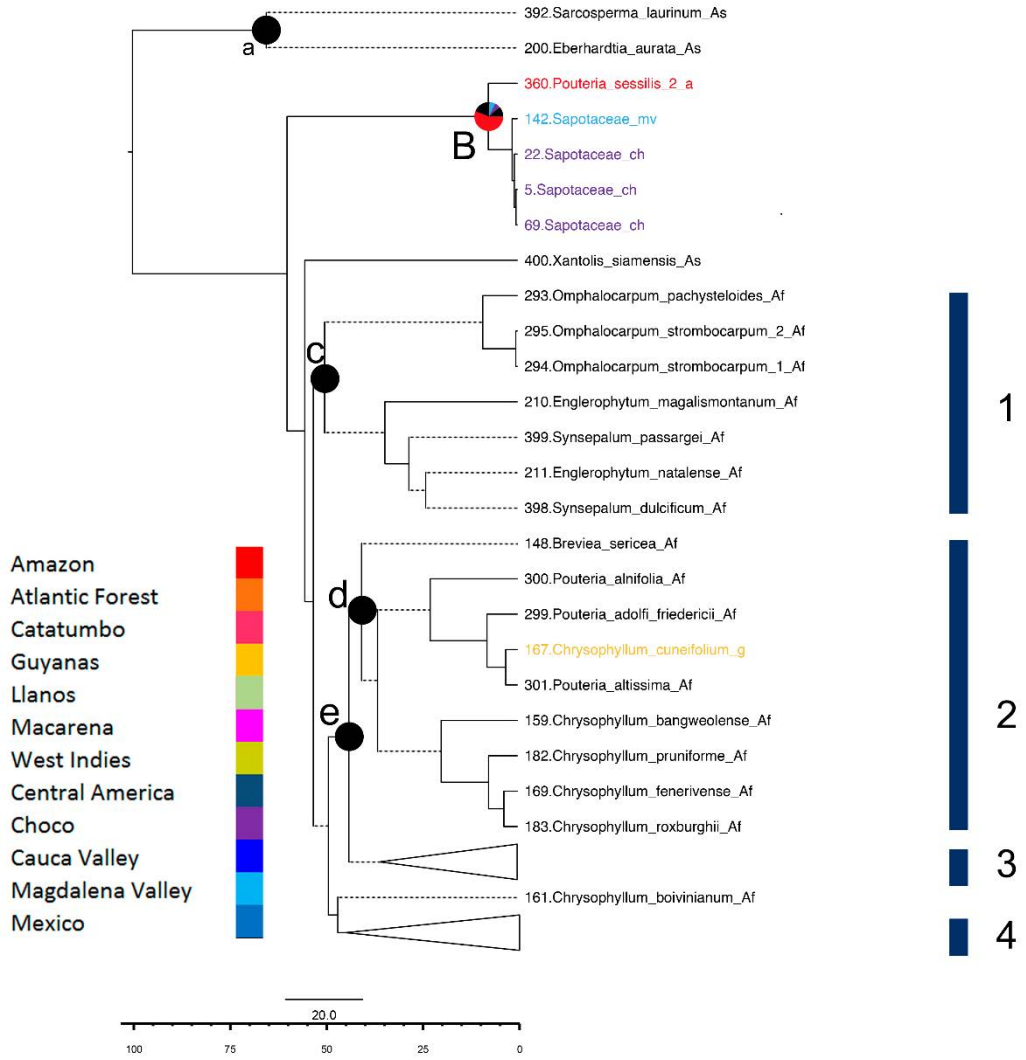
Figure 2.3. Maximum clade credibility tree from the BEAST analysis indicating relationships among Neotropical Sapotaceae lineages.
Three calibration points are indicated: node A – age estimate of Sapotaceae following Bremer *et al*., (2004); node B – Neotropical pollen; node C – *Malacantha alnifolia* pollen. Colours in tips and pie charts represent biogeographic areas in the Neotropics. These were chosen according to the presence of physical barriers that may have driven biotic distinction among areas (key in figure).

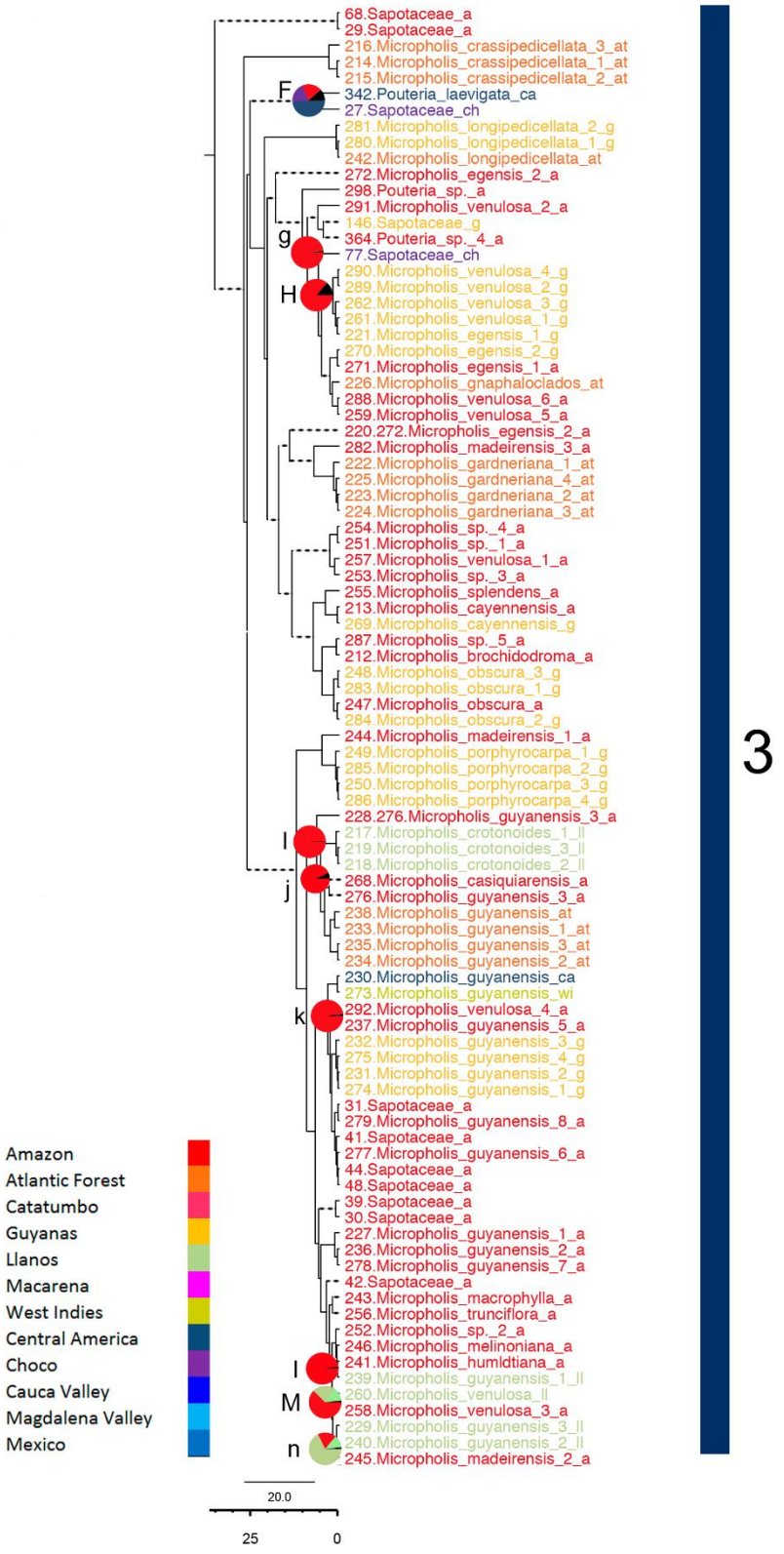
All figures in this chapter were produced using FigTree 1.4.2., ArcMap 10.1 and Photoshop CS6.

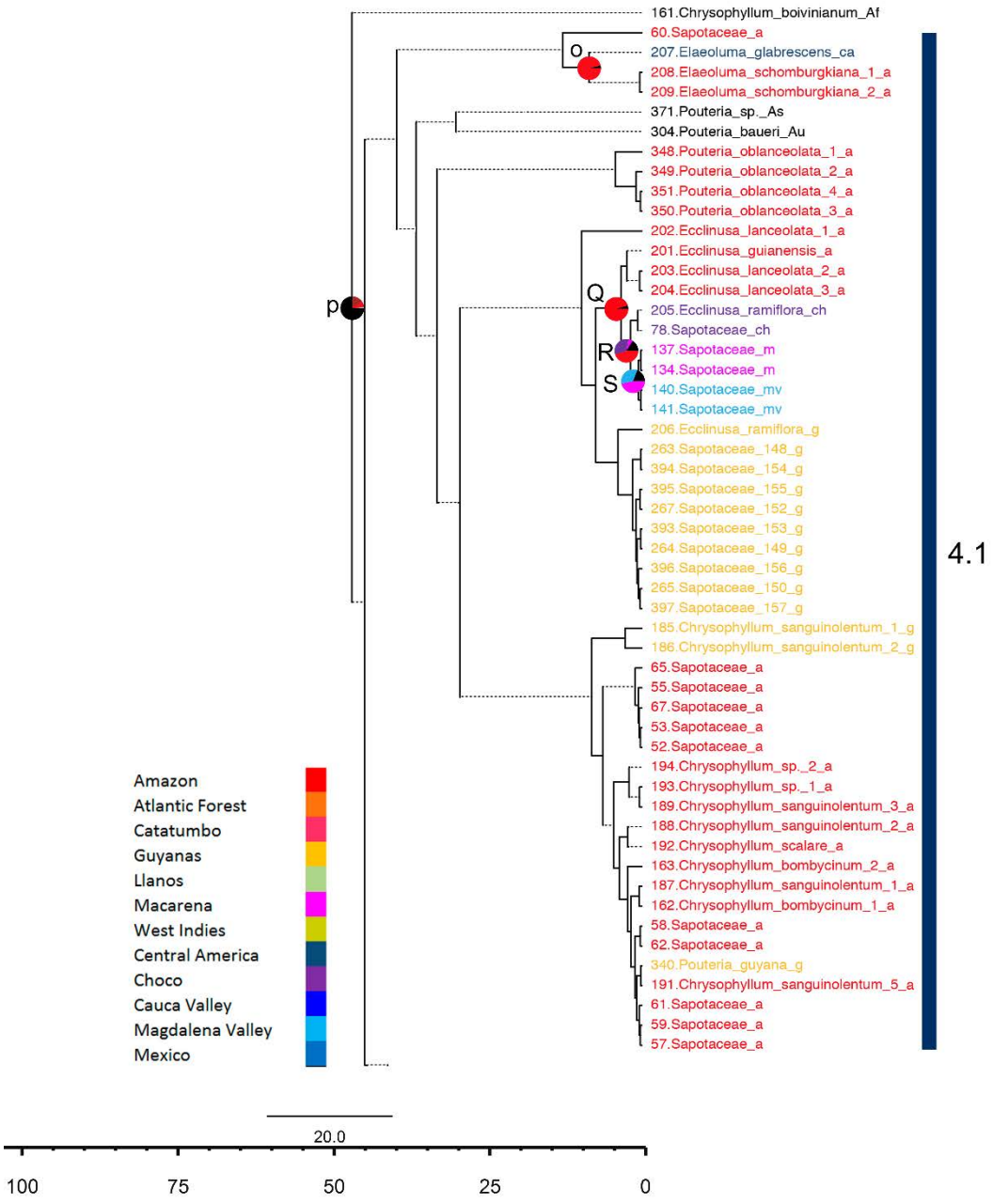
2.3 Results

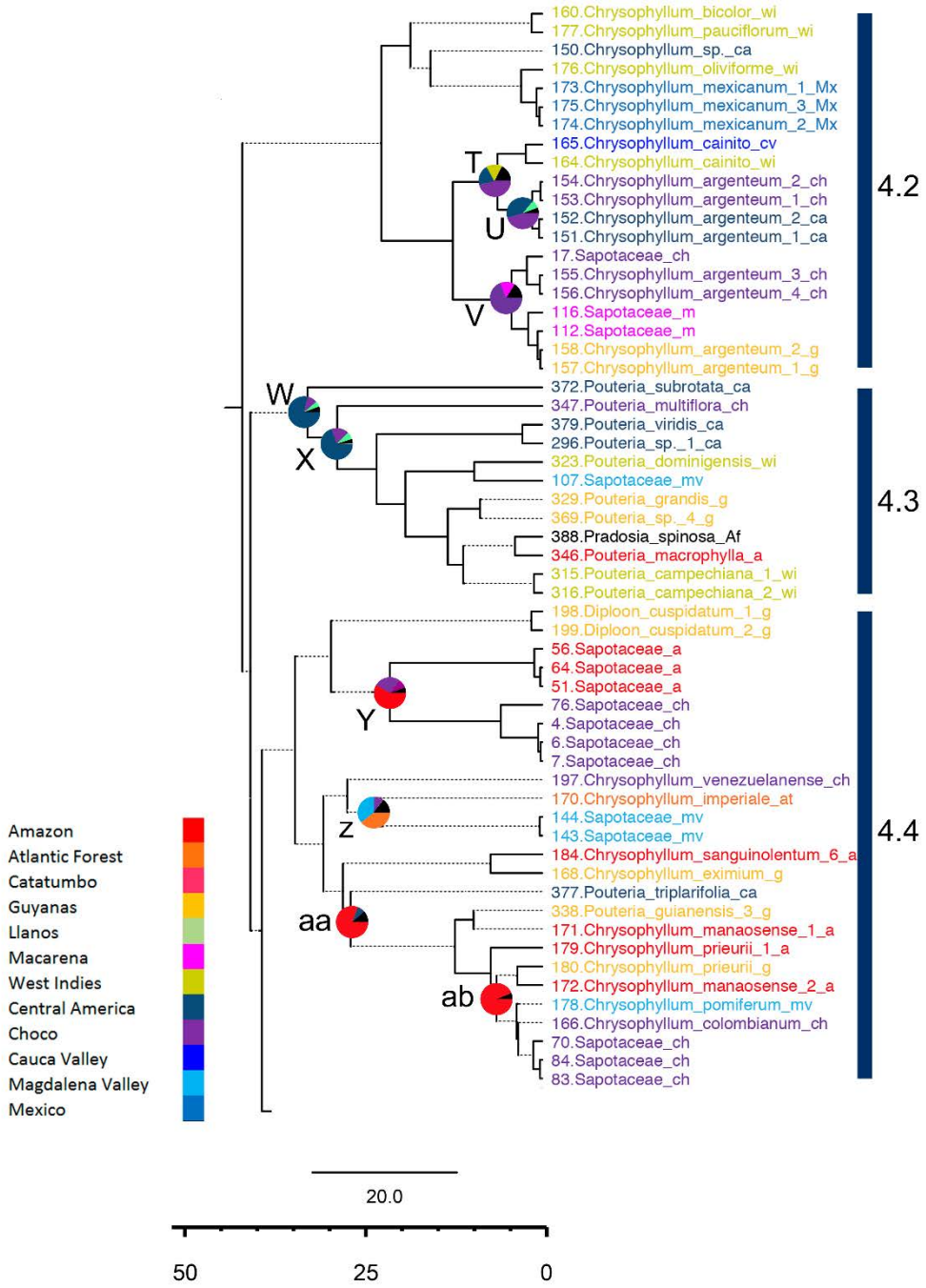
Thirty-six splits between west-Andean and east-Andean taxa were found from ca. 24 mya onwards, 17 with posterior probability (pp.) values higher than 0.9. Thirteen splits among lineages in Central America and South America were found from ca. 33 mya, seven of them with pp.> 0.9. Lastly, 13 splits in Sapotaceae occurred during the Pleistocene epoch, seven of them with pp. > 0.9. Each of the nodes that represent these splits are presented in Table 2.1 and Figure 2.4, with strongly supported nodes indicated with upper case letters and poorly supported ones in lower case.

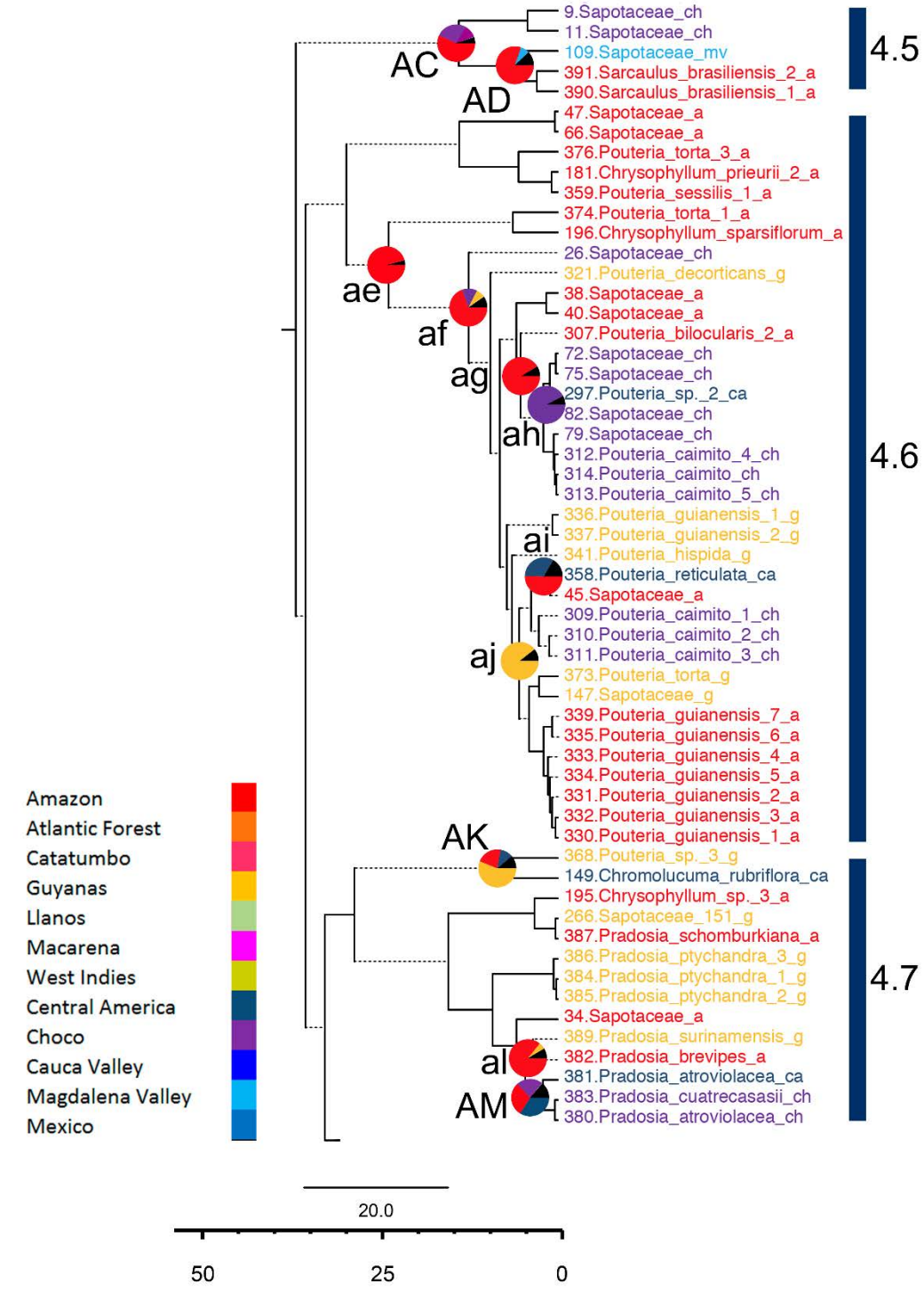
Node a in Figure 2.4 represents the earliest diversification of Asian Sapotaceae lineages at ca. 65 mya. The Asian Sapotaceae are not monophyletic with a well-supported neotropical clade nested within them and including taxa from the Amazon, Magdalena Valley and Choco (Node B Fig. 2.4. ca. 8.1 mya, pp. 1). The basal nodes of the phylogeny are weakly supported, but posterior probability values increase in younger clades. The crown node of African Sapotaceae lineages was found at ca. 53.46 mya. Two descendent African groups were found nested in the earliest clade. Those are indicated by Node c. (Fig. 2.4, pp. 0.34) diversifying at ca. 50 mya, and Node d diversifying at ca. 40 mya (Fig. 2.4, pp. 0.88).

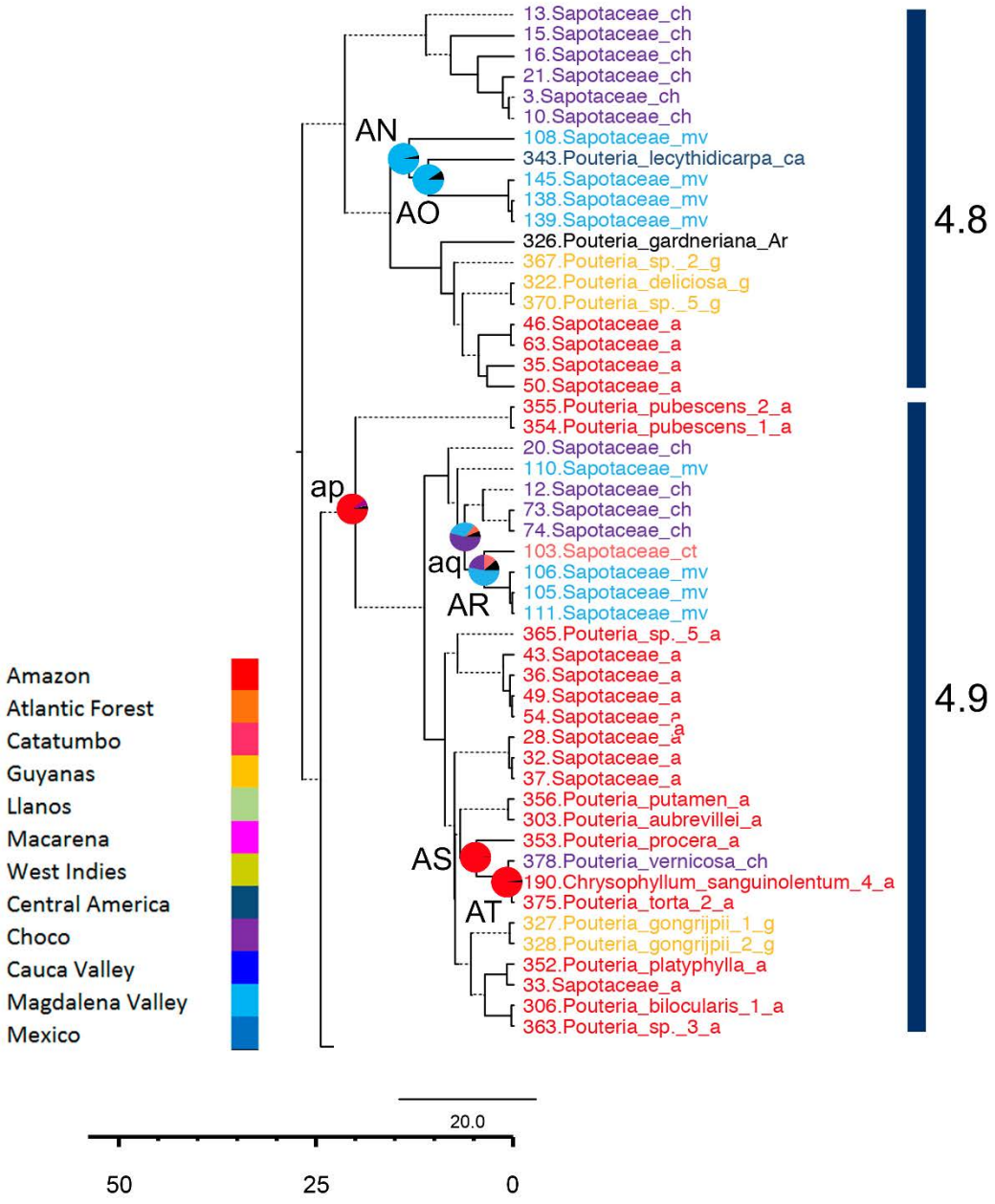

392.Sarcosperma_laurinum_As
200.Eberhardtia_aurata_As
360.Pouteria_sessilis_2_a
142.Sapotaceae_mv
22.Sapotaceae_ch
5.Sapotaceae_ch
69.Sapotaceae_ch
400.Xantolis_siamensis_As
293.Omphalocarpum_pachysteloides_Af
295.Omphalocarpum_strombocarpum_2_Af
294.Omphalocarpum_strombocarpum_1_Af
210.Englerophytum_magalismontanum_Af
399.Synsepalum_passargei_Af
211.Englerophytum_natalense_Af
398.Synsepalum_dulcificum_Af
148.Breviea_sericea_Af
300.Pouteria_alnifolia_Af
299.Pouteria_adolfi_friedericii_Af
167.Chrysophyllum_cuneifolium_g
301.Pouteria_altissima_Af
159.Chrysophyllum_bangweolense_Af
182.Chrysophyllum_pruniforme_Af
169.Chrysophyllum_fenerivense_Af
183.Chrysophyllum_roxburghii_Af
161.Chrysophyllum_boivinianum_Af
a
B
c
d
e
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
1
2
3
4
20.0
100
75
50
25
0

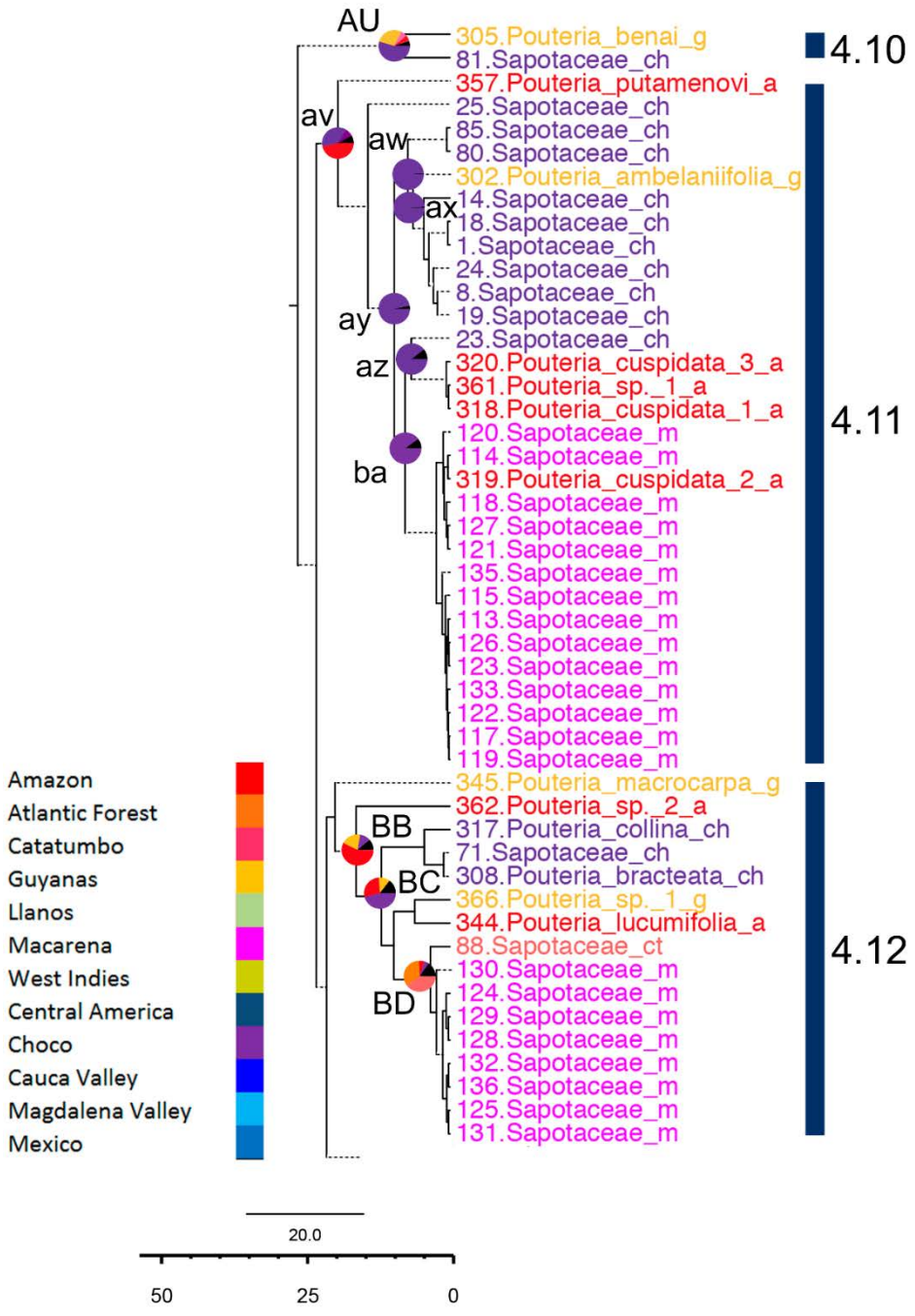

68.Sapotaceae_a
29.Sapotaceae_a
216.Micropholis_crassipedicellata_3_at
214.Micropholis_crassipedicellata_1_at
215.Micropholis_crassipedicellata_2_at
F
342.Pouteria_laevigata_ca
27.Sapotaceae_ch
281.Micropholis_longipedicellata_2_g
280.Micropholis_longipedicellata_1_g
242.Micropholis_longipedicellata_at
272.Micropholis_egensis_2_a
298.Pouteria_sp._a
291.Micropholis_venulosa_2_a
146.Sapotaceae_g
g
364.Pouteria_sp._4_a
77.Sapotaceae_ch
290.Micropholis_venulosa_4_g
289.Micropholis_venulosa_2_g
H
262.Micropholis_venulosa_3_g
261.Micropholis_venulosa_1_g
221.Micropholis_egensis_1_g
270.Micropholis_egensis_2_g
271.Micropholis_egensis_1_a
226.Micropholis_gnaphaloclados_at
288.Micropholis_venulosa_6_a
259.Micropholis_venulosa_5_a
220.272.Micropholis_egensis_2_a
282.Micropholis_madeirensis_3_a
222.Micropholis_gardneriana_1_at
225.Micropholis_gardneriana_4_at
223.Micropholis_gardneriana_2_at
224.Micropholis_gardneriana_3_at
254.Micropholis_sp._4_a
251.Micropholis_sp._1_a
257.Micropholis_venulosa_1_a
253.Micropholis_sp._3_a
255.Micropholis_splendens_a
213.Micropholis_cayennensis_a
269.Micropholis_cayennensis_g
287.Micropholis_sp._5_a
212.Micropholis_brochidodroma_a
248.Micropholis_obscura_3_g
283.Micropholis_obscura_1_g
247.Micropholis_obscura_a
284.Micropholis_obscura_2_g
244.Micropholis_madeirensis_1_a
249.Micropholis_porphyrocarpa_1_g
285.Micropholis_porphyrocarpa_2_g
250.Micropholis_porphyrocarpa_3_g
286.Micropholis_porphyrocarpa_4_g
228.276.Micropholis_guyanensis_3_a
217.Micropholis_crotonoides_1_II
l
219.Micropholis_crotonoides_3_II
218.Micropholis_crotonoides_2_II
268.Micropholis_casiquiarensis_a
j
276.Micropholis_guyanensis_3_a
238.Micropholis_guyanensis_at
233.Micropholis_guyanensis_1_at
235.Micropholis_guyanensis_3_at
234.Micropholis_guyanensis_2_at
230.Micropholis_guyanensis_ca
273.Micropholis_guyanensis_wi
292.Micropholis_venulosa_4_a
k
237.Micropholis_guyanensis_5_a
232.Micropholis_guyanensis_3_g
275.Micropholis_guyanensis_4_g
231.Micropholis_guyanensis_2_g
274.Micropholis_guyanensis_1_g
31.Sapotaceae_a
279.Micropholis_guyanensis_8_a
41.Sapotaceae_a
277.Micropholis_guyanensis_6_a
44.Sapotaceae_a
48.Sapotaceae_a
39.Sapotaceae_a
30.Sapotaceae_a
227.Micropholis_guyanensis_1_a
236.Micropholis_guyanensis_2_a
278.Micropholis_guyanensis_7_a
42.Sapotaceae_a
243.Micropholis_macrophylla_a
256.Micropholis_trunciflora_a
252.Micropholis_sp._2_a
246.Micropholis_melinoniana_a
l
241.Micropholis_humldtiana_a
239.Micropholis_guyanensis_1_II
M
260.Micropholis_venulosa_II
258.Micropholis_venulosa_3_a
229.Micropholis_guyanensis_3_II
n
240.Micropholis_guyanensis_2_II
245.Micropholis_madeirensis_2_a
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
20.0
25
0
3


161.Chrysophyllum_boivinianum_Af
60.Sapotaceae_a
207.Elaeoluma_glabrescens_ca
208.Elaeoluma_schomburgkiana_1_a
209.Elaeoluma_schomburgkiana_2_a
371.Pouteria_sp._As
304.Pouteria_baueri_Au
348.Pouteria_oblanceolata_1_a
349.Pouteria_oblanceolata_2_a
351.Pouteria_oblanceolata_4_a
350.Pouteria_oblanceolata_3_a
202.Ecclinusa_lanceolata_1_a
201.Ecclinusa_guianensis_a
203.Ecclinusa_lanceolata_2_a
204.Ecclinusa_lanceolata_3_a
205.Ecclinusa_ramiflora_ch
78.Sapotaceae_ch
137.Sapotaceae_m
134.Sapotaceae_m
140.Sapotaceae_mv
141.Sapotaceae_mv
206.Ecclinusa_ramiflora_g
263.Sapotaceae_148_g
394.Sapotaceae_154_g
395.Sapotaceae_155_g
267.Sapotaceae_152_g
393.Sapotaceae_153_g
264.Sapotaceae_149_g
396.Sapotaceae_156_g
265.Sapotaceae_150_g
397.Sapotaceae_157_g
185.Chrysophyllum_sanguinolentum_1_g
186.Chrysophyllum_sanguinolentum_2_g
65.Sapotaceae_a
55.Sapotaceae_a
67.Sapotaceae_a
53.Sapotaceae_a
52.Sapotaceae_a
194.Chrysophyllum_sp._2_a
193.Chrysophyllum_sp._1_a
189.Chrysophyllum_sanguinolentum_3_a
188.Chrysophyllum_sanguinolentum_2_a
192.Chrysophyllum_scalare_a
163.Chrysophyllum_bombycinum_2_a
187.Chrysophyllum_sanguinolentum_1_a
162.Chrysophyllum_bombycinum_1_a
58.Sapotaceae_a
62.Sapotaceae_a
340.Pouteria_guyana_g
191.Chrysophyllum_sanguinolentum_5_a
61.Sapotaceae_a
59.Sapotaceae_a
57.Sapotaceae_a
O
P
Q
R
S
4.1
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
20.0
100
75
50
25
0


160.Chrysophyllum_bicolor_wi
177.Chrysophyllum_pauciflorum_wi
150.Chrysophyllum_sp._ca
176.Chrysophyllum_oliviforme_wi
173.Chrysophyllum_mexicanum_1_Mx
175.Chrysophyllum_mexicanum_3_Mx
174.Chrysophyllum_mexicanum_2_Mx
165.Chrysophyllum_cainito_cv
164.Chrysophyllum_cainito_wi
154.Chrysophyllum_argenteum_2_ch
153.Chrysophyllum_argenteum_1_ch
152.Chrysophyllum_argenteum_2_ca
151.Chrysophyllum_argenteum_1_ca
17.Sapotaceae_ch
155.Chrysophyllum_argenteum_3_ch
156.Chrysophyllum_argenteum_4_ch
116.Sapotaceae_m
112.Sapotaceae_m
158.Chrysophyllum_argenteum_2_g
157.Chrysophyllum_argenteum_1_g
372.Pouteria_subrotata_ca
347.Pouteria_multiflora_ch
379.Pouteria_viridis_ca
296.Pouteria_sp._1_ca
323.Pouteria_dominigensis_wi
107.Sapotaceae_mv
329.Pouteria_grandis_g
369.Pouteria_sp._4_g
388.Pradosia_spinosa_Af
346.Pouteria_macrophylla_a
315.Pouteria_campechiana_1_wi
316.Pouteria_campechiana_2_wi
198.Diploon_cuspidatum_1_g
199.Diploon_cuspidatum_2_g
56.Sapotaceae_a
64.Sapotaceae_a
51.Sapotaceae_a
76.Sapotaceae_ch
4.Sapotaceae_ch
6.Sapotaceae_ch
7.Sapotaceae_ch
197.Chrysophyllum_venezuelanense_ch
170.Chrysophyllum_imperiale_at
144.Sapotaceae_mv
143.Sapotaceae_mv
184.Chrysophyllum_sanguinolentum_6_a
168.Chrysophyllum_eximium_g
377.Pouteria_triplarifolia_ca
338.Pouteria_guianensis_3_g
171.Chrysophyllum_manaosense_1_a
179.Chrysophyllum_prieurii_1_a
180.Chrysophyllum_prieurii_g
172.Chrysophyllum_manaosense_2_a
178.Chrysophyllum_pomiferum_mv
166.Chrysophyllum_colombianum_ch
70.Sapotaceae_ch
84.Sapotaceae_ch
83.Sapotaceae_ch
T
U
V
W
X
Y
Z
aa
ab
4.2
4.3
4.4
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
20.0
50
25
0


9.Sapotaceae_ch
11.Sapotaceae_ch
109.Sapotaceae_mv
391.Sarcaulus_brasiliensis_2_a
390.Sarcaulus_brasiliensis_1_a
47.Sapotaceae_a
66.Sapotaceae_a
376.Pouteria_torta_3_a
181.Chrysophyllum_prieurii_2_a
359.Pouteria_sessilis_1_a
374.Pouteria_torta_1_a
196.Chrysophyllum_sparsiflorum_a
26.Sapotaceae_ch
321.Pouteria_decorticans_g
38.Sapotaceae_a
40.Sapotaceae_a
307.Pouteria_bilocularis_2_a
72.Sapotaceae_ch
75.Sapotaceae_ch
297.Pouteria_sp._2_ca
82.Sapotaceae_ch
79.Sapotaceae_ch
312.Pouteria_caimito_4_ch
314.Pouteria_caimito_ch
313.Pouteria_caimito_5_ch
336.Pouteria_guianensis_1_g
337.Pouteria_guianensis_2_g
341.Pouteria_hispida_g
358.Pouteria_reticulata_ca
45.Sapotaceae_a
309.Pouteria_caimito_1_ch
310.Pouteria_caimito_2_ch
311.Pouteria_caimito_3_ch
373.Pouteria_torta_g
147.Sapotaceae_g
339.Pouteria_guianensis_7_a
335.Pouteria_guianensis_6_a
333.Pouteria_guianensis_4_a
334.Pouteria_guianensis_5_a
331.Pouteria_guianensis_2_a
332.Pouteria_guianensis_3_a
330.Pouteria_guianensis_1_a
368.Pouteria_sp._3_g
149.Chromolucuma_rubriflora_ca
195.Chrysophyllum_sp._3_a
266.Sapotaceae_151_g
387.Pradosia_schomburkiana_a
386.Pradosia_ptychandra_3_g
384.Pradosia_ptychandra_1_g
385.Pradosia_ptychandra_2_g
34.Sapotaceae_a
389.Pradosia_surinamensis_g
382.Pradosia_brevipes_a
381.Pradosia_atroviolacea_ca
383.Pradosia_cuatrecasasii_ch
380.Pradosia_atroviolacea_ch
AC
AD
ae
af
ag
ah
ai
aj
AK
al
AM
4.5
4.6
4.7
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
20.0
50
25
0


13.Sapotaceae_ch
15.Sapotaceae_ch
16.Sapotaceae_ch
21.Sapotaceae_ch
3.Sapotaceae_ch
10.Sapotaceae_ch
AN
108.Sapotaceae_mv
343.Pouteria_lecythidicarpa_ca
145.Sapotaceae_mv
AO
138.Sapotaceae_mv
139.Sapotaceae_mv
326.Pouteria_gardneriana_Ar
367.Pouteria_sp._2_g
322.Pouteria_deliciosa_g
370.Pouteria_sp._5_g
46.Sapotaceae_a
63.Sapotaceae_a
35.Sapotaceae_a
50.Sapotaceae_a
4.8
355.Pouteria_pubescens_2_a
354.Pouteria_pubescens_1_a
ap
20.Sapotaceae_ch
110.Sapotaceae_mv
12.Sapotaceae_ch
73.Sapotaceae_ch
74.Sapotaceae_ch
aq
103.Sapotaceae_ct
106.Sapotaceae_mv
105.Sapotaceae_mv
AR
111.Sapotaceae_mv
365.Pouteria_sp._5_a
43.Sapotaceae_a
36.Sapotaceae_a
49.Sapotaceae_a
54.Sapotaceae_a
28.Sapotaceae_a
32.Sapotaceae_a
37.Sapotaceae_a
356.Pouteria_putamen_a
303.Pouteria_aubrevillei_a
353.Pouteria_procera_a
AS
378.Pouteria_vernicosa_ch
190.Chrysophyllum_sanguinolentum_4_a
AT
375.Pouteria_torta_2_a
327.Pouteria_gongrijpii_1_g
328.Pouteria_gongrijpii_2_g
352.Pouteria_platyphylla_a
33.Sapotaceae_a
306.Pouteria_bilocularis_1_a
363.Pouteria_sp._3_a
4.9
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
20.0
50
25
0


AU
av
aw
ax
ay
az
ba
BB
BC
BD
305.Pouteria_benai_g
81.Sapotaceae_ch
357.Pouteria_putamenovi_a
25.Sapotaceae_ch
85.Sapotaceae_ch
80.Sapotaceae_ch
302.Pouteria_ambelaniifolia_g
14.Sapotaceae_ch
18.Sapotaceae_ch
1.Sapotaceae_ch
24.Sapotaceae_ch
8.Sapotaceae_ch
19.Sapotaceae_ch
23.Sapotaceae_ch
320.Pouteria_cuspidata_3_a
361.Pouteria_sp._1_a
318.Pouteria_cuspidata_1_a
120.Sapotaceae_m
114.Sapotaceae_m
319.Pouteria_cuspidata_2_a
118.Sapotaceae_m
127.Sapotaceae_m
121.Sapotaceae_m
135.Sapotaceae_m
115.Sapotaceae_m
113.Sapotaceae_m
126.Sapotaceae_m
123.Sapotaceae_m
133.Sapotaceae_m
122.Sapotaceae_m
117.Sapotaceae_m
119.Sapotaceae_m
345.Pouteria_macrocarpa_g
362.Pouteria_sp._2_a
317.Pouteria_collina_ch
71.Sapotaceae_ch
308.Pouteria_bracteata_ch
366.Pouteria_sp._1_g
344.Pouteria_lucumifolia_a
88.Sapotaceae_ct
130.Sapotaceae_m
124.Sapotaceae_m
129.Sapotaceae_m
128.Sapotaceae_m
132.Sapotaceae_m
136.Sapotaceae_m
125.Sapotaceae_m
131.Sapotaceae_m
4.10
4.11
4.12
Amazon
Atlantic Forest
Catatumbo
Guyanas
Llanos
Macarena
West Indies
Central America
Choco
Cauca Valley
Magdalena Valley
Mexico
20.0
50
25
0

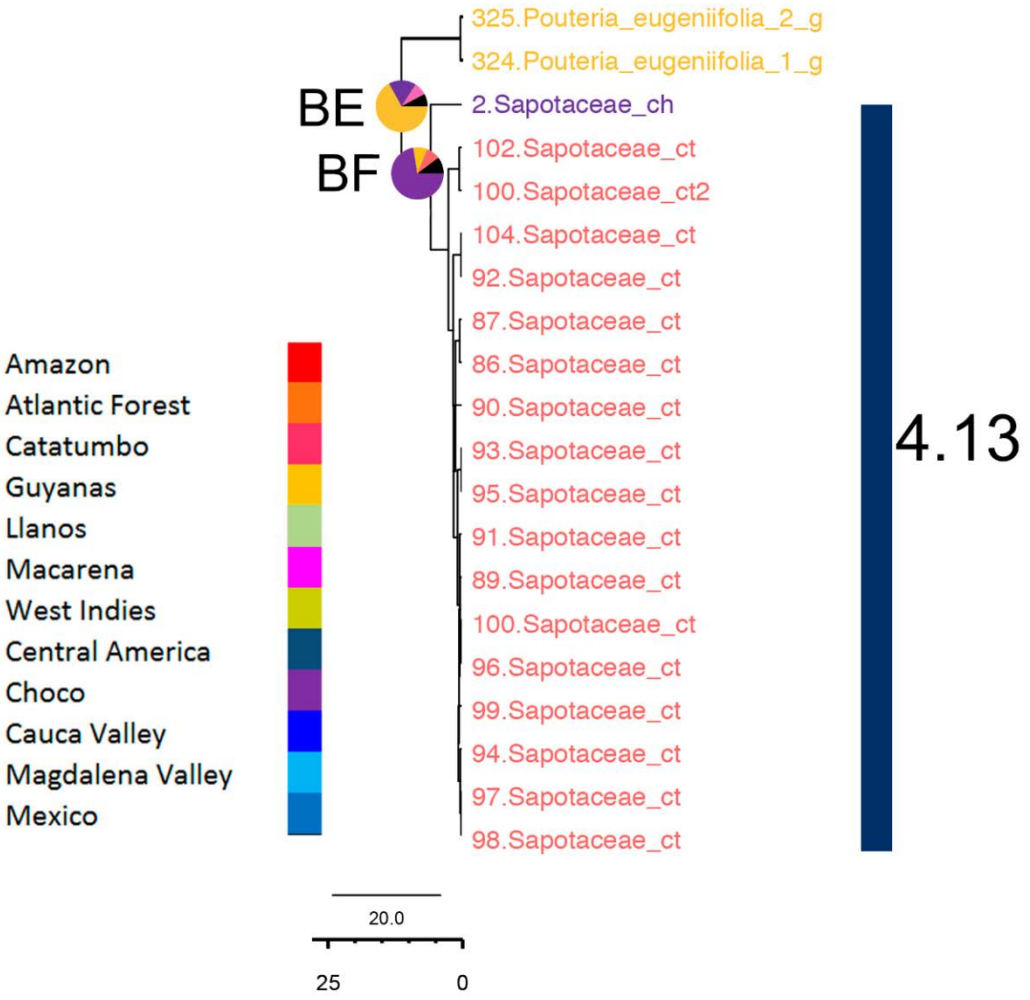


Figure 2.4. Maximum clade credibility tree from the BEAST analyses.
Letters represent splits across the Central American sea way (assuming a date of ca. 13 – 15 mya for the rise of the Panama Isthmus), splits among areas at the east of the Eastern Cordillera in Colombia, splits among areas at the west and the east of the Eastern Cordillera in Colombia, splits between Africa and the Neotropics, and splits among Africa and/or Asia. Capital letters are used for well supported nodes (pp.>0.9) and small letters for nodes with low support (pp.<0.9, dotted lines). Blue bars and numbers indicate division of clades among the phylogeny. Colours in tips and pie charts represent biogeographic areas in the Neotropics. These were chosen according to the presence of physical barriers that may have driven biotic distinction among areas. Areas are indicated at the ending of the tips labels: Amazon = a, Africa = Af, Argentina = Ar, Asia = As, Atlantic Forest = at, Australia = Au, Central America = ca, Choco = ch , Catatumbo = ct, Cauca Valley = cv, Guyanas = g, Llanos = ll, Macarena = m, Magdalena Valley = mv, México = Mx, West Indies = Wi.

We found two major neotropical clades (Clade 3 and Clade 4) derived from African lineages at ca. 44 mya (Node e. Fig. 2.4, pp. 0.17) and ca. 47 mya (Node p. Fig. 2.4, pp. 0.13), respectively. Node W represents a split at ca. 33 mya between Choco and Central America (pp. 0.99). In our analyses, this was the earliest well supported split between Choco and any other area in the Neotropics. Further splits between Choco and the Guyanas occurred at ca. 13 (Node af. Fig 2.4, pp. 0.86) 7.8 (Node aw. Fig 2.4, pp. 0.71) and 6.5 mya (Node BF. Fig 2.4, pp. 0.98. Well supported splits between Choco and Central America occurred at ca. 28 mya (Node X. Fig. 2.4, pp. 0.96), and more recently at 5.6 mya (Node F. Fig. 2.4, pp. 0.99), 2.7 mya (Node AM. Fig. 2.4, pp. 1), and 2 mya (Node U. Fig. 2.4, pp. 1).

Three well supported splits between Central America and other areas in South America occurred at ca. 15.9, 13 and 9.4 mya (Nodes AN, AO, AK. Fig. 2.4). Six additional splits among lineages in Central America and South America were found from 27.1 mya to 1.6 mya, but without strong statistical support (Nodes aa, o, ai, k, aj. Fig. 2.4).

The oldest split between lineages from Choco and the Amazon was found at ca. 24 mya (Node ae. Fig. 2.4, pp.0.87). However, this and a later split at ca. 23.7 mya (Node ap. Fig. 2.4.) are weakly supported (pp. 0.87 and 0.89, respectively). Node Y (Fig. 2.4.) is the earliest well supported split (pp. 0.99) between lineages from Choco and the Amazon at ca. 21.6 mya. Six additional well-supported splits between taxa in Choco and the Amazon were found at ca. 16 (Node BB. Fig. 2.4, pp. 1), 12.3 (Node BC. Fig. 2.4, pp. 0.99), 6 (Node AS. Fig. 2.4. pp. 0.99), 4 (Node Q. Fig. 2.4. Pp. 1) and 1.4 mya (Node AT. Fig. 2.4, pp. 1). Eight splits with pp.< 0.9 occurred between taxa in the Amazon and Choco from ca. 19.7 mya to ca. 5.1 mya (Nodes av, az, g, ay, ab, ag, and al. Fig. 2.4).

The oldest split between Choco and other areas in northern South America such as the Cauca valley was found at ca. 6.8 mya (node T. Fig. 2.4, pp. 1), and between Choco and Magdalena Valley at ca. 14.3 mya (node AC. Fig. 2.4, pp. 1). Another split between Choco and Magdalena Valley was dated to ca. 2.4 mya (e.g. Node R. Fig. 2.4, pp. 0.95).

Table 2.1. Summary of the divergence time estimation and ancestral state reconstruction analyses in the Neotropics.
Divergence time estimation analyses were carried out using BEAST v1.8.4 (Drummond and Rambaut, 2007), and ancestral state reconstruction analyses were performed using RASP (Yu *et al.*, 2015). HPD are given for those nodes with posterior probability values higher than 0.7. Nodes with posterior probability values >0.9 are represented by capital letters. Splits were divided into: Trans-Andean (between Choco/Cauca valley/Magdalena Valley and Amazon/Atlantic forest/Llanos/Macarena/Catatumbo); east-Andean (between Amazon, Catatumbo, Llanos and Macarena); Across the Panama Isthmus (between Central America/Mexico and South American areas); trans-Atlantic (between Paleotropics and Neotropics); and Paleotropical for areas in Africa and Asia.

Node	Posterior probability	Mean age and HPD (Mya)	Type of split	Ancestral Area (likelihood %)
a	0.83	65.66 [35.22, 94.92]	Trans-Atlantic	Asia (0.99)
aa	0.19	27.1	Panama Isthmus	Amazon (0.80)
ab	0.32	6.93	Trans-Andean	Amazon (0.93)
AC	1	14.3 [7.47, 20.29]	Trans-Andean	Amazon (0.44)
AD	1	6.11 [2.45, 8.92]	Trans-Andean	Amazon (0.80)
ae	0.87	24 [15.9, 31.37]	Trans-Andean	Amazon (0.66)
af	0.86	13.01 [6.99, 19.49]	Panama Isthmus	Amazon (0.69)
ag	0.25	5.8	Trans-Andean	Amazon (0.91)
ah	0.2	1.1	Panama Isthmus	Choco (0.92)
ai	0.23	1.6	Panama Isthmus	Amazon (0.50)
aj	0.39	6	Trans-Andean	Guyanas (0.89)
AK	1	9.4 [3.66, 14.79]	Panama Isthmus	Guyanas (0.56)
al	0.81	5.1 [2.30, 6.91]	Trans-Andean	Amazon (0.85)
AM	1	2.7 [0.69, 3.70]	Panama Isthmus	Central America (0.33)
AN	0.95	15.9 [9.64, 21.13]	Panama Isthmus	Magdalena valley (0.87)
AO	0.92	13.1 [7.24, 18.06]	Panama Isthmus	Magdalena valley (0.89)
ap	0.89	23.7 [15.95, 30.56]	Trans-Andean	Amazon (0.77)
aq	0.3	7.8	Trans-Andean	Choco (0.54)
AR	1	4.96 [1.90, 6.78]	Trans-Andean	Magdalena valley (0.52)
AS	0.99	6 [2.74, 8.41]	Trans-Andean	Amazon (0.97)
AT	1	1.4 [0.06, 1.76]	Trans-Andean	Amazon (0.97)
AU	1	10.65 [4.5, 16.12]	Panama Isthmus	Choco (0.54)
av	0.15	19.79	Trans-Andean	Amazon (0.46)
aw	0.71	7.84 [4.18, 10.31]	Trans-Andean	Choco (0.98)
ax	0.39	6.94	Trans-Andean	Choco (0.96)
ay	0.04	7.2	Trans-Andean	Choco (0.89)
az	0.35	10.16	Trans-Andean	Choco (0.96)

Node	Posterior probability	Mean age and HPD (Mya)	Type of split	Ancestral Area (likelihood %)
B	1	8.1 [2.89, 12.96]	Trans-Andean	Amazon (0.23)
ba	0.05	8.2	Trans-Andean	Choco (0.90)
BB	1	16 [11.09, 20.74]	Trans-Andean	Amazon (0.26)
BC	0.99	12.3 [7.74, 15.61]	Trans-Andean	Choco (0.17)
BD	1	3.9 [1.33, 5.48]	Trans-Andean	Catatumbo (0.39)
BE	0.96	12.1 [6.4, 17.07]	Trans-Andean	Guyanas (0.67)
BF	0.98	6.5 [2.83, 9.43]	Trans-Andean	Choco (0.72)
c	0.34	50	Paleotropical	Africa (0.94)
d	0.88	40 [35.08, 46.19]	Paleotropical	Africa (0.97)
e	0.17	44	Trans-Atlantic	Africa (0.56)
F	0.99	5.6 [1.11, 9.76]	Panama Isthmus	Central America (0.48)
g	0.41	9.74	Trans-Andean	Amazon (0.98)
H	0.98	6.57 [3.24, 8.07]	Trans-Andean	Amazon (0.86)
I	0.96	7 [3.95, 9.68]	East-Andean	Amazon (0.92)
j	0.12	5.9	East-Andean	Amazon (0.82)
k	0.03	3.8	Panama Isthmus	Amazon (0.97)
L	0.11	1	East-Andean	Amazon (0.98)
M	0.95	1.2 [0.01, 1.49]	East-Andean	Amazon (0.62)
n	0.1	1	East-Andean	Llanos (0.67)
o	0.48	9.09	Panama Isthmus	Amazon (0.96)
p	0.13	47	Trans-Atlantic	Africa (0.45)
Q	1	4 [1.49, 5.47]	Trans-Andean	Amazon (0.42)
R	0.95	2.4 [0.55, 3.17]	Trans-Andean	Amazon (0.20)
S	0.99	1.5 [0.05, 1.11]	Trans-Andean	Macarena (0.45)
T	1	6.8 [2.84, 9.64]	Trans-Andean	Choco (0.09)
U	1	2 [0.23, 2.75]	Panama Isthmus	Choco (0.43)
V	1	4.8 [1.79, 7]	Trans-Andean	Choco(0.65)
W	0.99	33 [24.18, 40.64]	Panama Isthmus	Central America (0.54)
X	0.96	28 [20.68, 36.33]	Panama Isthmus	Central America (0.62)
Y	0.99	21.6 [12.62, 29.78]	Trans-Andean	Amazon (0.56)
z	0.73	23 [12.68, 34.53]	Trans-Andean	Atlantic forest (0.39)

Node z at ca. 23 mya (pp. 0.73) indicates the time when lineages in the Magdalena valley and the Atlantic Forest in Brazil last shared a common ancestor. The earliest split between taxa in the Magdalena valley and the Amazon was found at ca. 8.1 mya (Node B. Fig. 2.4, pp. 1), taxa shared between these areas also diverged later at ca. 6.1 mya (Node AD. Fig. 2.4, pp. 1). The age of node S at ca. 1.5 mya (pp. 0.99) represents a younger split between lineages from the Magdalena valley and Macarena.

Two main groups of lineages from Macarena are represented by Node ba and Node BD. Node ba represents the most recent common ancestor shared between lineages from Macarena and Choco/Amazon at ca. 8.2 mya. This relationship is weakly supported (pp. 0.05). Node BD indicates the most recent common ancestor shared between taxa in Catatumbo and in Macarena at ca. 3.9 mya (pp. 1). Taxa in Macarena were found to also share a common ancestor with taxa in the Amazon at ca. 4 mya (Node Q. Fig. 2.4, pp.1) and in Choco at ca. 4.8 (Node V. Fig. 2.4, pp. 1) and 2.4 mya (Node R. Fig. 2.4, pp. 0.95).

The main Catatumbo clade (Clade 4.13) was found to diverge from lineages from Choco at ca. 6.5 mya (Node BF. Fig. 2.4, pp. 0.98). The stem node of that clade was found at ca. 12 (pp. 0.96) and represents the split between taxa in the Guyanas and clade 4.13. Taxa from Catatumbo split from taxa in the Magdalena valley at ca. 4.9 mya and 7.8 mya (Nodes AR and aq. Fig. 2.4, pp. 1 and 0.3). Lineages from Los Llanos split from taxa from the Amazon at ca. 1 mya (Node n. pp. 0.1 and Node L. pp. 0.11), 1.2 mya (Node M. pp. 0.95), 5.9 mya (Node j. pp. 0.12) and 7 mya (Node I. pp. 0.96).

2.4 Discussion

2.4.1 Arrival of Sapotaceae in the Neotropics

Sapotaceae are a pantropical family that according to our analyses first diverged in Asia at ca. 100 mya [HDP 93 – 105 mya], and that later colonised African lowland forest at ca. 53 mya. The Paleotropical groups are not monophyletic and we found a well-supported neotropical clade nested within the Asian species with a stem node age of ca. 60 mya [HDP 56 – 67 mya]. This is the first record for a possible Sapotaceae migration from Asia to the Neotropics, and it represents the oldest age of

divergence of Chrysophylloideae in these analyses. This estimate is younger than that of Bartish *et al.,* (2011), but both the age of Sapotaceae and of Chrysophylloideae in our data are consistent with Bremer *et al.,* (2004) and Armstrong *et al.,* (2014).

We found evidence for later migration to the Neotropics from Africa at ca. 47 and ca. 44 mya. These dispersal events postdate the isolation of South America from Africa, and occurred at the same time as the existence of the megathermal boreotropical rain forest in the northern hemisphere during the Palaeocene/Eocene (Wolfe, 1975). Sapotaceae could have colonised the Neotropics via the boreotropics as has been suggested based on dated molecular phylogenies of numerous other groups such as Annonaceae (Richardson *et al.,* 2004), Burseraceae (Weeks *et al.,* 2005), Malpighiaceae (Davis *et al.,* 2002), Melastomataceae, (Renner *et al.,* 2001), Meliaceae (Muellner *et al.,* 2006), Moraceae (Zerega *et al.,* 2005) and Rubiaceae (Antonelli *et al.,* 2009). Alternatively, colonization could have been by transatlantic long-distance dispersal, or perhaps via volcanic island hopping at the Rio Grande Rise and the Walvis Ridge (Parish, 1993), though the existence of islands in these locations is controversial (Pennington and Dick, 2004). According to the ancestral reconstruction in RASP, the earliest record of Sapotaceae in the Neotropics was found in the Amazon (Fig. 2.4), from where it later dispersed to other areas in the continent.

2.4.2 Diversification of Sapotaceae during the Andean uplift

Since arrival in the Neotropics, Sapotaceae diversified at a constant rate until ca. 20 mya. At ca. 20 mya there is a drop in the production of new lineages followed by an acceleration at ca. 10 mya (Fig. 2.5). It could be argued that this latter acceleration coincided with Andean uplift. The uplift of the Andes coincides with an acceleration in diversification rate that could have been caused by vicariance events. We acknowledge that the uplift of the Andes could have promoted diversification in the family in other ways, possibly by creating new habitats and by altering the edaphic and hydrologic systems in lowland areas (Burnham *et al.,* 1999; Hoorn *et al.,* 2010) but here we focus on the possible role of vicariance caused by the Andean orogeny.

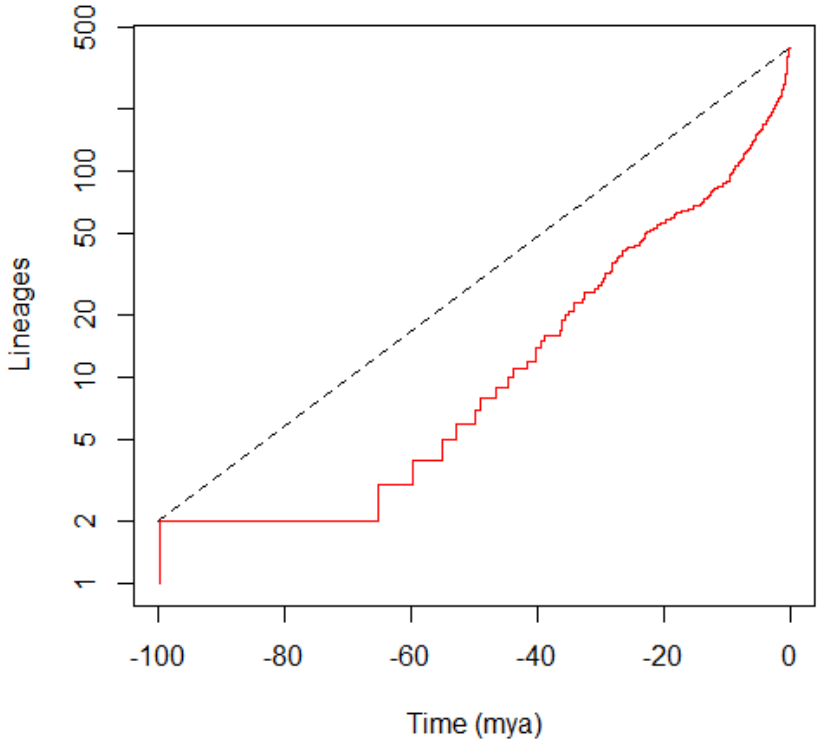


Figure 2.5. Lineages through time plot for neotropical Sapotaceae.
Rates of diversification in neotropical Sapotaceae species accelerated from ca. 10 mya onwards. This acceleration in diversification rates may have been caused by the Andean uplift as a result of isolation of taxa at either side of those mountains, and of the creation of new habitats and hydrologic systems in northern South America.

2.4.3 Trans-Andean splits

Thirty-six splits were found between taxa at either side of the Eastern Cordillera from ca. 24 mya onwards (Fig. 2.4 and 2.6). According to the hypothesis of Figure 2.2 we would expect splits on either side of the Eastern Cordillera to have occurred from c. 11.8 mya. In the analyses performed here, 27 (13 with pp.>0.9 and 14 with pp.<0.9) out of 36 splits between western and eastern Andean taxa occurred from ca. 11.8 mya onwards (Fig. 2.4 and 2.6). Timing of isolation across the Cordillera is spread through time rather than concentrated at one point as a vicariance model would predict. This indicates that isolation of Sapotaceae lineages on either side of the Andes after the mid Miocene, the time frame of major uplift in the northern South America, occurred only in some cases, and that migration across the Andean mountains has occurred among Sapotaceae lineages (Trenel *et al.*, 2007; Albert *et al.*, 2006; Piere *et al.*, 2006 and Arrivillaga *et al.*, 2002).

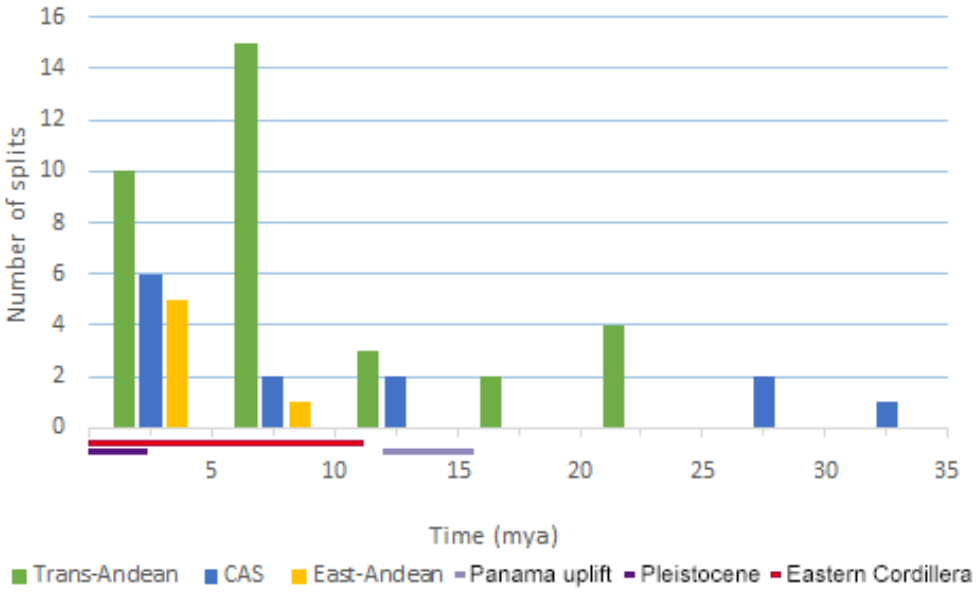


Figure 2.6. Number of phylogenetic splits found in Sapotaceae through geological time in northern South America.
Types of splits were divided into Trans Andean: between Choco/Cauca valley/Magdalena Valley and Amazon/Atlantic Forest/Llanos/Macarena/Catatumbo; east-Andean: between Amazon, Catatumbo, Llanos and Macarena; and Across CAS (Central American Seaway): between Central America/Mexico and South American areas.

Splits in Sapotaceae could have also been caused by the occurrence of seasonally dry biomes such as the seasonally dry tropical forests (SDTF) and savannahs. These areas may have prevented dispersal of wet forest restricted species resulting in allopatric speciation, in the same way as montane uplift, especially in lineages with a lack of tolerance to seasonality (Honorio *et al.,* 2014; Winterton *et al.,* 2014), or for taxa with inefficient dispersal mechanisms and not able to migrate through drier regions. This has been shown in Sapotaceae where a wet forest lineage of the genus *Manilkara* split on either side of the drier Cerrado and Caatinga biomes at a time consistent with diversification of taxa restricted to those more mesic ecosystems (Armstrong *et al.,* 2014; Simon *et al.,* 2009). If this were true for other lineages of Sapotaceae, splits between areas where the main barrier among patches of lowland rain forest are SDTF and savannahs (e.g. Dry areas in Huila and Arauca. Fig. 2.1) would be found in our phylogeny at the time when these biomes were formed.

Splits between the Macarena and Magdalena valley, areas that are separated by the Andes and to the north by the drier zones of Tatacoa and the Upper Magdalena Valley, occurred at node S (ca. 1.5 mya). Other Magdalena Valley-Amazonia splits

occurred at ca. 19 mya and ca. 6.11 mya. These splits may have occurred as a result of the formation of the dry biomes that currently lie between these wet forest areas. This potentially suggests an earliest age of ca. 19 mya [HDP 13-25 mya] for the formation of dry forest on the northern coast of Colombia, which coincides with both fossil and phylogenetic evidence for dry forest elsewhere in the Andes (Pennington *et al.,* 2004). Similarly, there is one split between Catatumbo and Macarena that occurred at node BD (ca. 3.9 mya). If we assume that migration by Sapotaceae was not possible across drier biomes and that no long-distance dispersal occurred, that age estimate could be indicative of the age of the formation of the dry forest and llanos (savannah) ecosystems that lie between Catatumbo and La Macarena/Amazonia (Simon *et al.,* 2009; Pennington and Hughes, 2014).

Splits between Choco and areas to the east of the Amazon occurred at nodes af (ca. 13 mya), aw (ca. 7.8 mya) and BF (ca. 6.5 mya). Barriers that may have caused those splits include the Andean mountains together with areas of dry forests. In other cases, lineages in Choco and other areas to the east of the Eastern Cordillera such as the Amazon have well supported divergence times from ca. 21 mya until ca. 1.4 mya. Additional splits between Choco and the eastern Andean forets of la Macarena were found at ca. 4.8 mya (Fig. 2.4). Some of these splits occurred well after montane habitats were created in the Eastern Cordillera at ca. 11.8 mya (Hoorn *et al.,* 1995) and after major uplift in the Santander massif at ca. 19 – 14 mya (Kroonenberg *et al.,* 1990), suggesting that dispersal was also a main driver in Sapotaceae evolution.

There are no obvious barriers to migration in northern Colombia to the west of the Eastern Cordillera because patches of lowland rain forest found to the north of the Central and Western Cordilleras are connected. Therefore, migration events between the inter-Andean valleys and Choco could have taken place via rain forest fragments located in the departments of Antioquia, Cordoba and Bolivar (Fig. 2.1; Gentry, 1982), and through mountain passes where altitudes in the Eastern Cordillera are lower (red bars Fig. 2.1). This is represented in our phylogeny at node AC (ca. 14.3 mya) and node R (ca. 2.4 mya) among lineages at the Choco and the Magdalena valley.

Catatumbo individuals appear more closely related to lineages to the west of the Eastern Cordillera than to lineages from areas where the Andes are not a

geographic barrier to the east of the Eastern Cordillera. This is suggested based on splits from lineages in Catatumbo and Magdalena Valley at node AR (ca. 4.9 mya) and node aq (ca. 7.8 mya), and from lineages in Catatumbo Choco at node BF (ca. 6.5 mya).

These findings, showing migration events occurring consistently through time, agree with some recent studies which have suggested that dispersal of rain forest trees may have been frequent among areas separated by the Andes (e.g., Fine *et al.,* 2014; Dexter *et al.,* 2017 but see Pirie *et al.,* 2006 and Winterton *et al.,* 2014), and with reports of floristic affinities between the lowland rain forest of la Serrania de los Motilones in the Catatumbo, and the Choco, Amazon and the Magdalena inter-Andean Valley in Colombia (e.g., Dueñas *et al.,* 2007; A. Avella, pers. comm., 18 February 2016). They also agree with the results presented in Chapter 1 (Biotic homogeneity of biogeographic units in the Neotropics: a test with Sapotaceae) and with patterns of diversity in seasonally dry tropical forest (SDTF). Banda *et al.,* (2016) suggested that the seasonally dry forest of the Norte de Santander department, the Valle del Cauca department and the inter-Andean valleys in Colombia (Fig. 2.1) have high floristic similarity and are part of the same floristic group. This connectivity may have been due to some of the possible terrestrial migration routes discussed above, but one cannot discount the possibility that some of it was due to longer-distance dispersal events.

2.4.4 Dispersal across the Panama Isthmus

Thirteen migration events, seven well supported, were found between Central American and South American areas (Fig. 2.6). These occurred from ca. 33 mya onwards (Node W, Fig 2.4.), prior to and after the closure of the Panama land bridge, assuming a date of closure of ca. 13 – 15 mya (Hoorn & Flantua 2015; Bacon *et al.,* 2015). Our results corroborate previous work such as that of Cody *et al.,* (2010) who compared plant and animal migration patterns across the isthmus, and found that dispersal of plant taxa was spread through the Cenozoic and often preceded the closure of the Isthmus of Panama (whether regarded as occurring at 3 mya or 13-15 mya). They are also similar to those of Bartish *et al.* (2011) who suggested that the closure of the Central American sea way was a possible route for migration, but not an essential element for the expansion of Sapotaceae's distributional range. Evidence of

migration in our results is also younger than the estimated age for the Greater Antilles and Avis ridge (GAARlandia) land bridge. According to MacPhee and Iturralde (2005), during the Oligocene/Eocene boundary (ca. 35-33 mya), for about 1 mya, GAARlandia connected the northwest of South America to what at present corresponds to the eastern part of Jamaica, possibly facilitating migration of biotas between South America and the rest of the Neotropics (but see Antonelli and Sanmartín, 2011). Considering that there is evidence for transatlantic dispersal in Sapotaceae (e.g. Bartish *et al.,* 2011), migration could have taken place from South America to Central and North America, directly across the relatively narrower seaway present before the rise of the Isthmus of Panama. It could also have occurred via the Greater Antilles, which emerged in the Cretaceous, reaching periods of major uplift in the Eocene (Graham, 2010) thereby allowing biotic exchange in the Neotropics since ca. 45 mya.

2.4.5 Pleistocene Diversification

Evidence for the effects of the Pleistocene climatic changes on species diversification in neotropical forests has been controversial in the past, and has been thought to have had an impact in only certain taxa. For instance, the review presented by Moritz *et al.,* (2000) suggested that speciation in fauna from the lowland rain forest of South America mostly predates the Pleistocene and was therefore not influenced by Pleistocene climates. According to Pennington *et al*. (2004), the same is true for seasonally dry tropical forest (SDTF) plant taxa in South America, but not for the rain forest species of *Ruprechtia* (Polygonaceae) or SDTF species in some Central American genera. In the past two decades, other examples suggesting origin of neotropical rain forest species during the Pleistocene have accumulated, for example in *Inga* and Meliaceae (see Richardson *et al.,* 2001 and Koenen *et al.,* 2015). Thirteen splits (seven with pp.>0.9) among South American Sapotaceae occurred during the Pleistocene (Fig. 2.6). Our results add to a building picture that some diversification events in neotropical trees have occurred recently, however, it does not appear to be an important upturn in diversification rates during the Pleistocene (Fig. 2.5 and 2.6 and Table 2.1). In conclusion, whilst Pleistocene dates for speciation in Sapotaceae are consistent in timing with a speciation model involving changes in rain forest cover, there is no evidence for Pleistocene speciation being responsible for the majority of

Sapotaceae species diversity as "refuge theory" would suggest (Haffer, 1969; Prance, 1973; Richardson, 2001).

2.4.6 Non-monophyletic species in the rain forest biome

In addition to assessing the effects of climatic changes and geographic barriers in Sapotaceae, by including multiple accessions of the same species in our data set, we were able to reveal patterns of evolution within and between species across Sapotaceae's geographical range. We identified non-monophyletic species with wider distributions, which may be the ancestors of species with narrower ranges that are nested within them. This pattern was found in species such as *Chrysophyllum cainito* that is nested within *C. argenteum, Micropholis casiquariensis* that is nested within *M. guyanensis*, and *Pouteria reticulata* and *P. caimito* that are nested within the more broadly distributed *P. guianensis*. Wide distributional ranges in lowland rain forest could be the result of dispersal events followed by successful colonisation in areas where empty spaces were created after periods of frequent disturbance, potentially by drought (Pennington and Lavin, 2016). In some cases, isolation may be sufficient to lead to species formation in peripheral populations, where strong selection or drift may produce morphologically distinct species. In the case of *P. caimito* and *C. cainito,* changes in morphology, sufficient to warrant description as distinct species, could have been the result of human selection aiming to harvest Sapotaceae species for the commercial value of its edible fruit.

2.5 Further work

Evidence for the effects of vicariance and dispersal on the distribution of tree species in the neotropical lowland rain forest biome was found using Sapotaceae as a model group. However, and even though this plant family is an important representative of the neotropical lowland rain forests, different groups may show evidence for different patterns. Therefore, further work aiming to add data of other representative taxa from the northern South American forests could provide important insights on the evolution of floras in the Neotropics. Such studies could continue adding samples from the lowland rain forest of northern South America, especially from countries like Colombia. Colombia is the second most diverse country in the world with more than 22,000 species of flowering plants of which more than 6,000 are endemics. This diversity is under-represented in scientific studies.

References

Albert, J. S., Lovejoy, N. R. & Crampton, W. G. R. 2006. Miocene tectonism and the separation of cis- and trans-Andean river basins: Evidence from neotropical fishes. Journal of South American Earth Sciences. 21: 14-27.

Anderberg, A. A. & Swenson, U. 2003. Evolutionary lineages in Sapotaceae (Ericales): A Cladistic analysis based on ndhf sequence data. International Journal of Plant Sciences. 164: 763-773.

Antonelli, A., Nylander, J. A. A., Persson, C. & Sanmartín, I. 2009. Tracing the impact of the Andean uplift on neotropical plant evolution. Proceedings of the National Academy of Sciences USA. 106: 9749-9754.

Antonelli, A. & Sanmartín, I. 2011. Why are there so many plant species in the Neotropics? Taxon. 60: 403-414.

Armstrong, K. E., Stone, G. N., Nicholls, J. A., Valderrama Escallón, E., Anderberg, A. A., Smedmark, J., … Richardson, J. E. 2014. Patterns of diversification amongst tropical regions compared: A case study in Sapotaceae. Frontiers in Genetics. 5: 1-13.

Arrivillaga, J. C., Norris, D. E., Feliciangeli, M. D., & Lanzaro, G. C. 2002. Phylogeography of the neotropical sand fly *Lutzomyia longipalpis* inferred from mitochondrial DNA sequences. Infection, Genetics and Evolution. 2: 83-95.

Bacon, C. D., Silvestro, D., Jaramillo C., Smith, B. T., Chakrabarty, P. & Antonelli, A. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proceedings of the National Academy of Sciences. 112: 6110-6115.

Banda, K., Delgado-Salinas, A., Dexter, K., Linares-Palomino, R., Oliveira-Filho, A., Prado, D., … Pennington, R. T. 2016. Plant diversity patterns in neotropical dry forests and their conservation implications. Science. 353: 1383-1387.

Bartish, I. V., Antonelli, A., Richardson, J. E. & Swenson, U. 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). Journal of Biogeography. 38: 177-190.

Bernal, R., Gradstein, S. R. & Celis, M. 2016. Catálogo de plantas y líquenes de Colombia. Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Instituto de Ciencias Naturales. 3060.

Bremer, B. 2009. Asterids. In the timetree of life (ed.) Hedges, S. B. & Kumar, S. Oxford. University Press. 177-187.

Bremer, K., Friis, E. & Bremer, B. 2004. Molecular phylogenetic dating of Asterids flowering plants shows early Cretaceous diversification. Systematic Biology. 53: 496-505.

Brumfield, R. T. & Edwards, S. V. 2007. Evolution into and out of the Andes: a Bayesian analysis of historical diversification in *Thamnophilus antshrikes*. Evolution. 61: 346-367.

Burnham, R. J. & Graham, A. 1999. The History of neotropical vegetation: New developments and status. Annals of the Missouri Botanical Garden. 86: 546-589.

Burnham, R. J. & Johnson, K. R. 2004. South American palaeobotany and the origins of neotropical rainforests. Philosophical Transactions of the Royal Society B. 359: 1595-1610.

Cody, S., Richardson, J. E., Rull. V., Ellis, C. & Pennington, R. T. 2010. The great American biotic interchange revisited. Ecography. 33: 1-7.

Colinvaux, P. A., Irion, G., Rasanen, M. E. & Bush, M. B. 2001. A paradigm to be discarded: Geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of amazonian speciation. Amazoniana. 16: 609-646.

Cortés-Ortiz, L., Bermingham, E., Rico, C., Rodriguez-Luna, E., Sampaio, I., & Ruiz-García, M. 2003. Molecular systematics and biogeography of the neotropical monkey genus, *Alouatta*. Molecular Phylogenetics and Evolution. 26: 64-81.

Darriba D, Taboada G. L., Doallo R. & Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods. 9: 772.

Davis, C. C., Bell, C. D., Mathews, S. & Donoghue, M. J. 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. Proceedings of the National Academy of Sciences USA. 99: 6833-6837.

Dexter, K. G., Lavin, M., Torke, B. M., Twyford, A. D., Kursar, T. A., Coley, P.D., … Pennington, R. T. 2017. Dispersal assembly of rain forest tree communities across the Amazon Basin. Proceedings of the National Academy of Sciences. 114: 2645-2650.

Drummond, A. J. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees BMC. Evolutionary Biology. 7: 214.

Dueñas, A., Betancur, J. & Galindo, R. 2007. Estructura y composición florística de un bosque húmedo tropical del Parque Nacional Natural Catatumbo Barí, Colombia. Colombia Forestal. 10: 26-39.

Duangjai, S., Samuel, R., Munzinger, J., Forest, F., Wallnofer, B., Barfuss, M. H. J., Fischer, G. & Chase, M. W. 2009. A multi-locus plastid phylogenetic analysis of the pantropical genus *Diospyros* (Ebenaceae), with an emphasis on the radiation and biogeographic origins of the New Caledonian endemic species. Molecular Phylogenetics and Evolution. 52: 602-620.

Farris, D. W., Jaramillo, C., Bayona, G., Restrepo-Moreno, S. A., Montes, C., Cardona, A., … Valencia, V. 2011. Fracturing of the Panamanian Isthmus during initial collision with South America. Geology. 39: 1007-1010.

Fine, P. V. A., Zapata, F. & Daly, D. C. 2014. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). Evolution. 68: 1988-2004.

Gentry, A. H. 1982. Neotropical floristic diversity: Phytogeographical connections between Central and South America: Pleistocene climatic fluctuations, or an accident of the Andean orogeny? Annals of the Missouri Botanical Garden. 69: 557-593.

Gonzalez, M. A., Eberhard, J. R., Lovette, I. J., Olson, S. L., & Bermingham, E. 2003. Mitochondrial DNA Phylogeography of the Bay Wren (Troglodytidae: *Thryothorus nigricapillus*) Complex. The Condor. 105: 228-238.

Gonzalez, M. A., Baraloto, C., Engel, J., Mori, S. A., Pétronelli, P., Riéra, B., Roger, A., Thébaud, C. & Chave, J. 2009. Identification of Amazonian trees with DNA barcodes. PLoS One. 4: e7483.

Graham, A. 2010. A natural history of the New World: the ecology and evolution of plants in the Americas. University of Chicago Press. 387.

Graham, A. 2011. The age and diversification of terrestrial new world ecosystems through Cretaceous and Cenozoic time. American Journal of Botany. 98: 336-351.

Gregory-Wodzicki, K. M. 2000. Uplift history of the central and northern Andes: A review. Geological Society of America Bulletin. 112: 1091-1105.

Guindon, S., Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology. 52: 696-704.

Haffer, J. 1969. Speciation in amazonian forest birds. Science. 165: 131-137.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98.

Ho, S. Y. W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. Journal of Avian Biology. 38: 409-414.

Ho, S. Y. W. & Phillips, M. J. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology. 58: 367-380.

Honorio Coronado, E. N., Dexter, K. G., Poelchau, M.F., Hollingsworth, P. M., Phillips, O. L. & Pennington, R.T. 2014. *Ficus insipida* subsp. insipida (Moraceae) reveals the role of ecology in the phylogeography of widespread neotropical rain forest tree species. Journal of Biogeography. 41: 1697-1709.

Hoorn, C., Guerrero J., Sarmiento, G. A. & Lorente, M. A. 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. Geology. 23: 237-240.

Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., … Antonelli, A. 2010. Amazonia through time: andean uplift, climate change, landscape evolution and biodiversity. Science. 330: 927-931.

Hoorn, C. & Flantua, S. 2015. An early start for the Panama land bridge. Science. 348: 186-187.

Huelsenbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17: 754-755.

IDEAM, IGAC, IAvH, Invemar, I. Sinchi e IIAP. 2007. Ecosistemas continentales, costeros y marinos de Colombia. Instituto de Hidrología, Meteorología y Estudios Ambientales, Instituto Geográfico Agustín Codazzi, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Instituto de Investigaciones Ambientales del Pacífico Jhon von Neumann, Instituto de Investigaciones Marinas y Costeras José Benito Vives De Andréis e Instituto Amazónico de Investigaciones Científicas Sinchi. Bogotá, D. C, 276 p. + 37 hojas cartográficas.

Jan du Chêne, R. E., Onyike, M. S. & Sowunmi, M. A. 1978. Some new Eocene pollen of the Ogwashi-Asabe Formation, south-eastern Nigeria. Revista Española de Micropaleontologia. 10: 285-322.

Kearse M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28: 1647-1649.

Koenen, E. J. M., Clarkson, J. J., Pennington, T. D. & Chatrou, L. W. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. New Phytologist. 207: 327-339.

Kroonenberg, S., Bakker, J. & Van der Wiel, M. 1990. Late Cenozoic uplift and paleogeography of the Colombian Andes: constraints on the development of high-andean biota. Geologie. 69: 279-290.

Macphee, R. D. E. & Iturralde-Vinent, M. 2005. The Interpretation of Caribbean Paleogeography: Reply to Hedges. In Alcover, J.A. & Bover, P. (ed.): Proceedings of the International Symposium "Insular Vertebrate Evolution: the Palaeontological Approach". Monografies de la Societat d'Història Natural de les Balears. 12: 175-184.

Miller, M. A., Pfeiffer, W. & Schwartz, T. 2010. "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA. 1-8.

Mora, A., Baby, P., Roddaz, M., Parra, M., Brusset, S., Hermoza, W. & Espurt, N. 2010. Tectonic history of the Andes and sub-Andean zones: implications for the development of the Amazon drainage basin. In: Hoorn, C. & Wesselingh, F. P. (ed.): Amazonia, Landscape and Species Evolution: A Look into the Past. Wiley-Blackwell Publishing, Reino Unido. 38-60.

Moritz, C., Patton, J. L., Schneider, C. J. & Smith, T. B. 2000. Diversification of rainforest faunas: An integrated molecular approach. Annual Review of Ecology and Systematics. 31: 533-563.

Morrone, J. 2001. Biogeografia de America Latina y el Caribe. First ed. Zaragoza: CYTED, ORCYT-UNESCO, Sociedad Entomológica Aragonesa (SEA).

Muellner, A. N., Savolainen, V., Samuel, R. & Chase, M.W. 2006. The mahogany family ‘‘out-of-Africa’’: divergence time estimation, global biogeographic patterns inferred from plastid rbcL DNA sequences, extant, and fossil distribution of diversity. Molecular Phylogenetics and Evolution. 40: 236-250.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., … Morrison, J.C. 2001. Terrestrial ecoregions of the world: a new map of life on earth. Bioscience. 51: 933-938.

Parish, J. T. 1993. The paleogeography of the opening South Atlantic. In The Africa-South America connection (ed.) George, W. & Lavocat, R. Oxford: Clarendon. 8-41.

Pennington. T. D. 1991. The genera of Sapotaceae. Royal Botanic Gardens Kew, New York Botanical Garden. 295.

Pennington, R. T., Lavin, M., Prado, D. E., Pendry, C.A., Pell, S.K. & Butterworth, C.A. 2004. Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences. 359: 515-537.

Pennington, R. T. & Dick, C. W. 2004. The role of immigrants in the assembly of the South American rainforest tree flora. Philosophical Transactions of the Royal Society B: Biological Sciences. 359: 1611-1622.

Pennington R. T. & Hughes, C. E. 2014. The remarkable congruence of New and Old World savannah origins. New Phytologist. 204: 4-6.

Pennington, R. T. & Lavin, M. 2016. The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. New Phytologist. 210: 25-37.

Pirie, M. D., Chatrou, L. W., Mols, J. B., Erkens, R. H. J. & Oosterhof, J. 2006. 'Andean-centred' genera in the short-branch clade of Annonaceae: Testing biogeographic hypotheses using phylogeny reconstruction and molecular dating. Journal of Biogeography. 33: 31-46.

Prance, G. T. 1973. Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon Basin, based upon evidence from distribution pattern in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae. Acta Amazon. 3: 5-28.

Rambaut, A. 2016. FigTree [computer software]. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Renner, S. 2004. Plant Dispersal across the tropical Atlantic by wind and sea currents. International Journal of Plant Sciences. 165: 23-33.

Renner, S.S., Clausing, G. & Meyer, K. 2001. Historical biogeography of Melastomataceae: the roles of Tertiary migration and long-distance dispersal. American Journal of Botany. 88: 1290-1300.

Reynel, C., Pennington, T. & Särkinen, T. 2013. Cómo se formó la diversidad ecológica del Perú. Lima, Perú. 412.

Richardson, J. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science. 293: 2242-2245.

Richardson, J. E., Chatrou, L. W., Mols, J. B., Erkens, R. H. J. & Pirie, M. D. 2004. Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. Philosophical Transactions of the Royal Society B: Biological Sciences. 359: 1495-1508.

Richardson, J. E., Bakar, A. M., Tosh, J., Armstrong, K., Smedmark, J., Anderberg, A. A., … Wilkie, P. 2014. The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae). Botanical Journal of the Linnean Society. 174: 130-140.

Ronquist, F. & Huelsenbeck, J. P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572-1574.

Rull, V. 1997. Sequence analysis of Western Venezuelan Cretaceous to Eocene sediments using palynology. Chrono-palaeoenvironmental and palaeovegetational approaches. Palynology. 21: 79-90.

Rull, V. 2000. Ecostratigraphic study of Paleocene and early Eocene palynological cyclicity in northern South America. Palaios. 15: 14-24.

Sánchez-C., D., Richardson, J. E., Hart, M., Serrano, J., Cárdenas, D. & Cortés-B., R. 2017 [Forthcoming]. DNA barcoding as a tool for identification of species of the genus Micropholis (Griseb.) Pierre (Sapotaceae). DNA Barcodes.

Simon, M. F., Grether, R., Queiroz, L. P. De, Skema, C., Pennington, R. T. & Hughes, C. E. 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by *in situ* evolution of adaptations to fire. 106: 48.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30: 1312-1313.

Swenson, U., Richardson, J. E. & Bartish, I. V. 2008. Multi-gene phylogeny of the pantropical subfamily Chrysophylloideae (Sapotaceae): Evidence of generic polyphyly and extensive morphological homoplasy. Cladistics. 24: 1006-1031.

Trénel, P., Gustafsson, M. H. G., Baker, W. J., Asmussen-Lange, C. B., Dransfield, J. & Borchsenius, F. 2007. Mid-Tertiary dispersal, not gondwanan vicariance explains distribution patterns in the wax palm subfamily (Ceroxyloideae: Arecaceae). Molecular Phylogenetics and Evolution. 45: 272-288.

Weeks, A., Daly, D. C. & Simpson, B. B. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. Molecular Phylogenetics and Evolution. 35: 85-101.

Whinnett, A., Zimmermann, M., Willmott, K. R., Herrera, N., Mallarino, R., Simpson, … Mallet, J. 2005. Strikingly variable divergence times inferred across an Amazonian butterfly 'suture zone'. Proceedings of the Royal Society, London, B. 272: 2525-2533.

Winship, T. 1990. Paleobiogeographic relationships of angiosperms from the Cretaceous and Early Tertiary of the North American area. Botanical Review. 56: 279-417.

Winterton, C., Richardson, J. E., Hollingsworth, M., Clark, A., Zamora, N., & Pennington, R. T. 2014. "Historical biogeography of the neotropical legume Genus *Dussia*: the Andes, the Panama Isthmus and the Chocó," in Paleobotany and Biogeography: A Festschrift for Alan Graham in his 80th Year (ed.) Stevens, W. D., Montiel, O. M. & Raven, P. St Louis, MO: Missouri Botanical Press. 389- 404.

Wolfe, J. A. 1975. Some Aspects of Plant Geography of the northern Hemisphere During the Late Cretaceous and Tertiary. Annals of the Missouri Botanical Garden. 62: 264-279.

Wortel, R. 1984. Spatial and temporal variations in the Andean subduction zone. Journal of the Geological Society. 141: 783-791.

Yu, Y., Harris A. J., Blair C. & He, X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. Molecular Phylogenetics and Evolution. 87: 46-49.

Zerega, N. J. C., Clement, W. L., Datwyler, S. L. & Weiblen, G.D. 2005. Biogeography and diversification times in the mulberry family (Moraceae). Molecular Phylogenetics and Evolution. 37: 402-416.

Appendices

Appendix 2.1

Taxon	Colector/genebank	Locality
Breviea sericea (A. Chev.) Aubrév. & Pellegr.	Letouzey 8319 (P) DQ246666	Africa
Chromolucuma rubriflora Ducke	Anderberg *et al*. 20 (S)	Central America
Chrysophyllum argenteum Jacq.	Baraloto 3203 (CAY) NH200191	Guyanas
Chrysophyllum argenteum Jacq.	Baraloto, Engel & Riera s.n. (CAY) NH200307	Guyanas
Chrysophyllum argenteum Jacq.	Dick 79 (STRI)	Central America
Chrysophyllum argenteum Jacq.	Dick 80 (STRI)	Central America
Chrysophyllum argenteum Jacq.	Dick 81 (STRI)	Choco
Chrysophyllum argenteum Jacq.	Dick 82 (STRI)	Choco
Chrysophyllum argenteum Jacq.	Dick 85 (STRI)	Choco
Chrysophyllum argenteum Jacq.	Dick 87 (STRI)	Choco
Chrysophyllum bangweolense R.E. Fr.	Malaisse 9600 (WAG)	Africa
Chrysophyllum bicolor Poir.	JF912987	West Indies
Chrysophyllum boivinianum (Pierre) Baehni	McPherson 14426 (MO, P, WAG)	Africa
Chrysophyllum bombycinum T.D. Penn.	Garcia 8671 (E)	Amazon
Chrysophyllum bombycinum T.D. Penn.	Garcia 8673 (E)	Amazon
Chrysophyllum cainito L	Chantaranothai 2304 (KKU)	West Indies
Chrysophyllum cainito L	Richardson *et al*. 357 (E)	Cauca Valley
Chrysophyllum colombianum (Aubrév.) T.D. Penn.	Gentry, Faber–Langendoen & Monsalve 56806 (K)	Choco
Chrysophyllum cuneifolium (Rudge) A. DC.	Mori *et al*. 19135 (K)	Guyanas
Chrysophyllum eximium Ducke	Baraloto 3135 (CAY)	Guyanas
Chrysophyllum fenerivense (Aubrév.) G.E. Schatz & L. Gaut	NH200106	Africa
Chrysophyllum imperiale (Linden ex K. Koch & Fintelm.) Benth. & Hook. f.	Pennington s.n. (S)	Atlantic Forest

Taxon	Colector/genebank	Locality
Chrysophyllum L.	Anderberg *et al*. 31	Central America
Chrysophyllum L.	Garcia 8581 (E)	Amazon
Chrysophyllum L.	Garcia 8647 (E)	Amazon
Chrysophyllum L.	TBC	Amazon
Chrysophyllum manaosense (Aubrév.) T.D. Penn.	Richardson *et al*. 334 (E)	Amazon
Chrysophyllum manaosense (Aubrév.) T.D. Penn.	TBC	Amazon
Chrysophyllum mexicanum Brandegee ex Standl.	JP256 JF912991	Mexico
Chrysophyllum mexicanum Brandegee ex Standl.	JP341 JF912989	Mexico
Chrysophyllum mexicanum Brandegee ex Standl.	JP398 JF912990	Mexico
Chrysophyllum oliviforme L.	Guttiérrez & Nilsson 1 (S)	West Indies
Chrysophyllum pauciflorum Lam	JF912986	West Indies
Chrysophyllum pomiferum (Eyma) T.D. Penn.	Pennington *et al*. 12386	Magdalena Valley
Chrysophyllum prieurii A. DC.	Baraloto, Engler & Chave s.n. (CAY) P01860227	Guyanas
Chrysophyllum prieurii A. DC.	Garcia 8583 (E)	Amazon
Chrysophyllum prieurii A. DC.	Richardson *et al*. 326 (E)	Amazon
Chrysophyllum pruniforme Pierre ex Engl.	Jongkind 3762 (WAG)	Africa
Chrysophyllum roxburghii G. Don	Randrianasolo 33 (WAG)	Africa
Chrysophyllum sanguinolentum (Pierre) Baehni	Baraloto 3016 (CAY)	Guyanas
Chrysophyllum sanguinolentum (Pierre) Baehni	Baraloto 3196 (CAY) NH200177	Guyanas
Chrysophyllum sanguinolentum (Pierre) Baehni	FJ037869	Amazon
Chrysophyllum sanguinolentum (Pierre) Baehni	Garcia 8236 (E)	Amazon
Chrysophyllum sanguinolentum (Pierre) Baehni	Garcia 8645 (E)	Amazon
Chrysophyllum sanguinolentum (Pierre) Baehni	Richardson *et al*. 305 (E)	Amazon

Taxon	Colector/genebank	Locality
Chrysophyllum sanguinolentum (Pierre) Baehni	Richardson *et al*. 310 (E)	Amazon
Chrysophyllum sanguinolentum (Pierre) Baehni	Richardson *et al*. 348 (E)	Amazon
Chrysophyllum scalare T.D. Penn.	Richardson *et al*. 320 (E)	Amazon
Chrysophyllum sparsiflorum Klotzsch ex Miq.	DQ021882	Amazon
Chrysophyllum venezuelanense (Pierre) T.D. Penn	Ståhl *et al*. 5755 (S)	Choco
Diploon cuspidatum (Hoehne) Cronquist	Baraloto 3209 (CAY) NL110264	Guyanas
Diploon cuspidatum (Hoehne) Cronquist	Pennington *et al*. 13843 (U)	Guyanas
Eberhardtia aurata (Pierre ex Dubard) Lecomte	Hao 534 (S) EF558617	Asia
Ecclinusa guianensis Eyma	Ducke Res. 05-906 (K)	Amazon
Ecclinusa lanceolata (Mart. & Eichler) Pierre	Garcia 8675 (E)	Amazon
Ecclinusa lanceolata (Mart. & Eichler) Pierre	Garcia 8689 (E)	Amazon
Ecclinusa lanceolata (Mart. & Eichler) Pierre	Richardson *et al*. 314 (E)	Amazon
Ecclinusa ramiflora Mart.	Irwing *et al*. 55081 (S)	Guyanas
Ecclinusa ramiflora Mart.	Monsalve B1319 (K)	Choco
Elaeoluma glabrescens (Mart. & Eichler) Aubrév.	Anderberg *et al*. 33 (S)	Central America
Elaeoluma schomburgkiana (Miq.) Baill.	DQ246679	Amazon
Elaeoluma schomburgkiana (Miq.) Baill.	Keel & Koelho 243 (S)	Amazon
Englerophytum magalismontanum (Sond.) T.D. Penn.	Swenson & Karis 631 (S) DQ246680	Africa
Englerophytum natalense (Sond.) T.D. Penn.	Kayombo 3483 (S) AY552150	Africa
Micropholis (Griseb.) Pierre	Alvarez 737	Amazon
Micropholis (Griseb.) Pierre	Alvarez 799	Amazon
Micropholis (Griseb.) Pierre	Avila *et al*. 2733	Amazon
Micropholis (Griseb.) Pierre	Garcia 8618 (E)	Amazon
Micropholis (Griseb.) Pierre	Sanchez *et al*. 22	Amazon
Micropholis brochidodroma T.D. Penn.	JR-JH-1	Amazon
Micropholis casiquiarensis Aubrév.	Richardson *et al*. 340 (E)	Amazon

Taxon	Colector/genebank	Locality
Micropholis cayennensis T.D. Penn	Baraloto, Engel & Riera s.n. (CAY) NH200222	Guyanas
Micropholis cayennensis T.D. Penn	FJ037875.1	Amazon
Micropholis crassipedicellata (Mart. & Eichler ex Miq.) Pierre	JQ434164.1	Atlantic Forest
Micropholis crassipedicellata (Mart. & Eichler ex Miq.) Pierre	JQ434165.1	Atlantic Forest
Micropholis crassipedicellata (Mart. & Eichler ex Miq.) Pierre	NH200222	Atlantic Forest
Micropholis crotonoides (Pierre) Pierre	ESR89	Llanos
Micropholis crotonoides (Pierre) Pierre	JRM24	Llanos
Micropholis crotonoides (Pierre) Pierre	Ramirez *et al*. 50	Llanos
Micropholis egensis (A. DC.) Pierre	Avila 2746	Amazon
Micropholis egensis (A. DC.) Pierre	Betancur, Echeverry, Kress & Roesel 2848 (K)	Amazon
Micropholis egensis (A. DC.) Pierre	Dionizia, Coêlho & Ernesto 73 (U)	Amazon
Micropholis egensis (A. DC.) Pierre	DQ246681	Guyanas
Micropholis egensis (A. DC.) Pierre	FJ037876.1	Guyanas
Micropholis gardneriana (A. DC.) Pierre	JQ434158.1	Atlantic Forest
Micropholis gardneriana (A. DC.) Pierre	JQ434158.2	Atlantic Forest
Micropholis gardneriana (A. DC.) Pierre	JQ434159.1	Atlantic Forest
Micropholis gardneriana (A. DC.) Pierre	JQ434160.1	Atlantic Forest
Micropholis gardneriana (A. DC.) Pierre	JQ434161.1	Atlantic Forest
Micropholis guyanensis (A. DC.) Pierre	Avila 2684	Amazon
Micropholis guyanensis (A. DC.) Pierre	Baraloto 3107 (CAY) NL110516	Guyanas
Micropholis guyanensis (A. DC.) Pierre	Baraloto, Engel & Riviera s.n. (CAY) NH200477	Guyanas
Micropholis guyanensis (A. DC.) Pierre	DQ246682.1	Central America
Micropholis guyanensis (A. DC.) Pierre	FJ037878.1	Guyanas
Micropholis guyanensis (A. DC.) Pierre	Garcia 8242 (E)	Amazon
Micropholis guyanensis (A. DC.) Pierre	Garcia 8620 (E)	Amazon

Taxon	Colector/genebank	Locality
Micropholis guyanensis (A. DC.) Pierre	JQ434166.1	Atlantic Forest
Micropholis guyanensis (A. DC.) Pierre	JQ434167.1	Atlantic Forest
Micropholis guyanensis (A. DC.) Pierre	JQ434168.1	Atlantic Forest
Micropholis guyanensis (A. DC.) Pierre	KF943856.1	Atlantic Forest
Micropholis guyanensis (A. DC.) Pierre	Miranda *et al*. 93	Llanos
Micropholis guyanensis (A. DC.) Pierre	Miranda *et al*. 93	Llanos
Micropholis guyanensis (A. DC.) Pierre	Pineros *et al*. 14	Amazon
Micropholis guyanensis (A. DC.) Pierre	Richardson	Amazon
Micropholis guyanensis (A. DC.) Pierre	Richardson *et al*. 328 (E)	Amazon
Micropholis guyanensis (A. DC.) Pierre	Taylor 11691 (MO)	West Indies
Micropholis guyanensis (A. DC.) Pierre	CR194	Llanos
Micropholis guyanensis (A. DC.) Pierre	JR-PA-08	Amazon
Micropholis guyanensis (A. DC.) Pierre	TBC	Amazon
Micropholis guyanensis (A. DC.) Pierre	TBC	Guyanas
Micropholis humboldtiana (Roem. & Schult.) T.D. Penn.	PU72 (COAH)	Amazon
Micropholis longipedicellata Aubrév.	Baraloto 3210 (CAY) FJ037879	Guyanas
Micropholis longipedicellata Aubrév.	Baraloto 3210 (CAY) NL110459	Guyanas
Micropholis longipedicellata Aubrév.	FJ037879.1	Atlantic Forest
Micropholis macrophylla (Krause) T.D. Penn.	EP102	Amazon
Micropholis madeirensis (Baehni) Aubrév.	Cortes *et al*. 2869	Amazon
Micropholis madeirensis (Baehni) Aubrév.	MS3660	Amazon
Micropholis madeirensis (Baehni) Aubrév.	Richardson *et al*. 329 (E)	Amazon
Micropholis melinoniana Pierre	Alvarez 1999	Amazon
Micropholis obscura T.D. Penn.	Baraloto 3217 (CAY) P00610293	Guyanas
Micropholis obscura T.D. Penn.	Baraloto 3218 (CAY) NH200217	Guyanas
Micropholis obscura T.D. Penn.	FJ037880.1	Guyanas

Taxon	Colector/genebank	Locality
Micropholis obscura T.D. Penn.	TBC	Amazon
Micropholis porphyrocarpa (Baehni) Monach.	Baraloto 3219 (CAY) NH200696	Guyanas
Micropholis porphyrocarpa (Baehni) Monach.	Baraloto 3220 (CAY) NH200404	Guyanas
Micropholis porphyrocarpa (Baehni) Monach.	FJ037882.1	Guyanas
Micropholis porphyrocarpa (Baehni) Monach.	FJ037883.1	Guyanas
Micropholis splendens Gilly ex Aubrév.	Avila *et al*. 2693	Amazon
Micropholis trunciflora Ducke	Cardenas 12266	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	AD8893	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	Assunção 122 (U)	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	Avila *et al*. 2704	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	Baraloto 3114 (CAY) NL110297	Guyanas
Micropholis venulosa (Mart. & Eichler) Pierre	Baraloto 3128 (CAY) NL110260	Guyanas
Micropholis venulosa (Mart. & Eichler) Pierre	DQ246683.1	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	FJ037884.1	Guyanas
Micropholis venulosa (Mart. & Eichler) Pierre	FJ037885.1	Guyanas
Micropholis venulosa (Mart. & Eichler) Pierre	Garcia 8005 (E)	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	Garcia 8238 (E)	Amazon
Micropholis venulosa (Mart. & Eichler) Pierre	Sanchez *et al*. 124	Llanos
Omphalocarpum pachysteloides Mildbr. ex Hutch. & Dalziel	Jongkind 2351 (WAG) AY552151	Africa
Omphalocarpum strombocarpum Y.B. Harv. & Lovett	Frimodt-Møller *et al*. 538 (c) DQ246685	Africa
Omphalocarpum strombocarpum Y.B. Harv. & Lovett	Frimodt-Møller *et al*. 538 (C)	Africa
Pouteria adolfi-friedericii (Engl.) A. Meeuse	Friis *et al*. 3502 (UPS)	Africa
Pouteria alnifolia (Baker) Roberty	Jongkind & Noyes 1322 (MO)	Africa

Taxon	Colector/genebank	Locality
Pouteria altissima (A. Chev.) Baehni	Friis *et al*. 4145 (UPS)	Africa
Pouteria ambelaniifolia (Sandwith) T.D. Penn.	Baraloto s.n. (CAY) NH200025	Guyanas
Pouteria Aubl.	TBC	Guyanas
Pouteria Aubl.	Anderberg *et al*. 51 (S)	Central America
Pouteria Aubl.	Anderberg *et al*. 60 (S)	Central America
Pouteria Aubl.	Armstrong 317	Asia
Pouteria Aubl.	Garcia 8623 (E)	Amazon
Pouteria Aubl.	Garcia 8632 (E)	Amazon
Pouteria Aubl.	Garcia 8640 (E)	Amazon
Pouteria Aubl.	Garcia 8641 (E)	Amazon
Pouteria Aubl.	Garcia 8777 (E)	Amazon
Pouteria Aubl.	Sabatier & Molino K383	Guyanas
Pouteria Aubl.	Sabatier & Molino K989	Guyanas
Pouteria Aubl.	Sabatier & Molino P213	Guyanas
Pouteria Aubl.	Sabatier & Molino Q423	Guyanas
Pouteria Aubl.	Sabatier & Molino V819	Guyanas
Pouteria Aubl.	TBC	
Pouteria aubrevillei Bernardi	Richardson *et al*. 327 (E)	Amazon
Pouteria baueri (Montrouz.) Baehni	TBC	Australia
Pouteria benai (Aubrév. & Pellegr.) T.D. Penn.	Mori *et al*. 27353 (K)	Guyanas
Pouteria bilocularis (H.J.P. Winkl.) Baehni	Garcia 8763 (E)	Amazon
Pouteria bilocularis (H.J.P. Winkl.) Baehni	Richardson *et al*. 345 (E)	Amazon
Pouteria bracteata T.D. Penn.	Palacios 13699 (K)	Choco
Pouteria caimito (Ruiz & Pav.) Radlk.	Richardson *et al*. 377 (E)	Choco
Pouteria caimito (Ruiz & Pav.) Radlk.	Richardson *et al*. 386 (E)	Choco
Pouteria caimito (Ruiz & Pav.) Radlk.	Richardson *et al*. 387 (E)	Choco

Taxon	Colector/genebank	Locality
Pouteria caimito (Ruiz & Pav.) Radlk.	Richardson *et al*. 388 (E)	Choco
Pouteria caimito (Ruiz & Pav.) Radlk.	Richardson *et al*. 390 (E)	Choco
Pouteria caimito (Ruiz & Pav.) Radlk.	Richardson *et al*. 391 (E)	Choco
Pouteria campechiana (Kunth) Baehni	Fagerlind 2754 (S)	West Indies
Pouteria campechiana (Kunth) Baehni	Wang W00798 (HAST)	West Indies
Pouteria collina (Little) T.D. Penn.	Gentry, Faber–Langendoen & Echevarria 62908 (K)	Choco
Pouteria cuspidata (A. DC.) Baehni	Garcia 8624 (E)	Amazon
Pouteria cuspidata (A. DC.) Baehni	Garcia 8769 (E)	Amazon
Pouteria cuspidata (A. DC.) Baehni	Richardson *et al*. 337 (E)	Amazon
Pouteria decorticans T.D. Penn.	Baraloto, Engel, Chave & Riera s.n. (CAY) NL110158	Guyanas
Pouteria deliciosa T.D. Penn.	Sabatier & Molino 903	Guyanas
Pouteria dominigensis (C.F. Gaertn.) Baehni	Gutiérrez & Nilsson 13 (S)	West Indies
Pouteria eugeniifolia (Pierre) Baehni	Baraloto 3136 (CAY) NH200045	Guyanas
Pouteria eugeniifolia (Pierre) Baehni	Baraloto, Engel, Chave & Riera s.n. (CAY) NH200157	Guyanas
Pouteria gardneriana (A. DC.) Radlk.	Schwarz 8216 (UPS)	Argentina
Pouteria gongrijpii Eyma	Baraloto 3087 (CAY) NL110210	Guyanas
Pouteria gongrijpii Eyma	Baraloto 3098 (CAY) NL110282	Guyanas
Pouteria grandis Eyma	Sabatier & Molino 995	Guyanas
Pouteria guianensis Aubl.	Baraloto & Engel s.n. (CAY) M17116516	Guyanas
Pouteria guianensis Aubl.	Baraloto, Engel, Chave & Riera 3102 (CAY) NL110300	Guyanas
Pouteria guianensis Aubl.	Garcia 8631 (E)	Amazon
Pouteria guianensis Aubl.	Poncy 1745 (P)	Guyanas
Pouteria guianensis Aubl.	Richardson *et al*. 309 (E)	Amazon
Pouteria guianensis Aubl.	Richardson *et al*. 323 (E)	Amazon

Taxon	Colector/genebank	Locality
Pouteria guianensis Aubl.	Richardson *et al*. 332 (E)	Amazon
Pouteria guianensis Aubl.	Richardson *et al*. 333 (E)	Amazon
Pouteria guianensis Aubl.	Richardson *et al*. 342 (E)	Amazon
Pouteria guianensis Aubl.	Richardson *et al*. 351 (E)	Amazon
Pouteria hispida Eyma	Mori *et al*. 25432 (NY)	Guyanas
Pouteria laevigata (Mart.) Radlk.	Anderberg *et al*. 49 (S)	Central America
Pouteria lecythidicarpa P.E. Sánchez & Poveda	Anderberg *et al*. 34 (S)	Central America
Pouteria lucumifolia (Reissek ex Maxim.) T.D. Penn.	Richardson *et al*. 306 (E)	Amazon
Pouteria macrocarpa (Mart.) D. Dietr.	Baraloto, Engel, Chave & Riera 3095 (CAY) NL110257	Guyanas
Pouteria macrophylla (Lam.) Eyma	Seidel *et al*. 5905 (K)	Amazon
Pouteria multiflora (A. DC.) Eyma	Villa & Rivaz 257 (BM)	Choco
Pouteria oblanceolata Pires	Garcia 8625 (E)	Amazon
Pouteria oblanceolata Pires	Garcia 8634 (E)	Amazon
Pouteria oblanceolata Pires	Garcia 8635 (E)	Amazon
Pouteria oblanceolata Pires	Richardson *et al*. 307 (E)	Amazon
Pouteria platyphylla (A.C. Sm.) Baehni	Richardson *et al*. 313 (E)	Amazon
Pouteria procera (Mart.) K. Hammer	Richardson *et al*. 339 (E)	Amazon
Pouteria pubescens (Aubrév. & Pellegr.) T.D. Penn.	Garcia 8636 (E)	Amazon
Pouteria pubescens (Aubrév. & Pellegr.) T.D. Penn.	TBC	Amazon
Pouteria putamen-ovi T.D. Penn.	Garcia 8774 (E)	Amazon
Pouteria putamen-ovi T.D. Penn.	TBC	Amazon
Pouteria reticulata (Engl.) Eyma	Anderberg *et al*. 7 (S)	Central America
Pouteria sessilis T.D. Penn.	Garcia 8637 (E)	Amazon
Pouteria sessilis T.D. Penn.	Richardson *et al*. 321 (E)	Amazon
Pouteria subrotata Cronquist	Anderberg *et al*. 48 (S)	Central America

Taxon	Colector/genebank	Locality
Pouteria torta (Mart.) Radlk.	Anderberg *et al*. 47 (S)	Amazon
Pouteria torta (Mart.) Radlk.	Baraloto, Engel & Riera (CAY) NH200224	Guyanas
Pouteria torta (Mart.) Radlk.	Richardson *et al*. 336 (E)	Amazon
Pouteria torta (Mart.) Radlk.	Richardson *et al*. 338 (E)	Amazon
Pouteria triplarifolia Standl. & L.O. Williams ex T.D. Penn	Villa *et al*. 1304 (BM)	Central America
Pouteria vernicosa T.D. Penn.	Swenson *et al*. 738 (S)	Choco
Pouteria viridis (Pittier) Cronquist	Faber–Langendoen & Renteria 1127 (K)	Central America
Pradosia atroviolacea Ducke	Anderberg *et al*. 52 (S)	Choco
Pradosia atroviolacea Ducke	Lindeman 6743 (U)	Central America
Pradosia brevipes (Pierre) T.D. Penn.	Cuatrecasas 13988 (K)	Amazon
Pradosia cuatrecasasii (Aubrév.) T.D. Penn.	Baraloto, Engel & Riera s.n. (CAY) NH200016	Choco
Pradosia ptychandra (Eyma) T.D. Penn.	Baraloto, Engel, Chave & Riera (CAY) NL110310	Guyanas
Pradosia ptychandra (Eyma) T.D. Penn.	Baraloto, Engel, Chave & Riera 3111 (CAY) NL110514	Guyanas
Pradosia ptychandra (Eyma) T.D. Penn.	Ducke Reserve 05-1829 (K)	Guyanas
Pradosia schomburgkiana (A. DC.) Cronquist	Isembe 508 (WAG)	Amazon
Pradosia spinosa Ewango & Breteler	Harris 1076 (U)	Africa
Pradosia surinamensis (Eyma) T.D. Penn.	Harris 1076 (U)	Guyanas
Sapotaceae Juss.	Guy 13	Guyanas
Sapotaceae Juss.	Guy 2	Guyanas
Sapotaceae Juss.	Guy 24	Guyanas
Sapotaceae Juss.	Guy 26	Guyanas
Sapotaceae Juss.	Guy 3	Guyanas
Sapotaceae Juss.	Guy 32	Guyanas
Sapotaceae Juss.	Guy 37	Guyanas
Sapotaceae Juss.	Guy 47	Guyanas
Sapotaceae Juss.	Guy 55	Guyanas
Sapotaceae Juss.	Guy 74	Guyanas
Sapotaceae Juss.	Guy 76	Guyanas

Taxon	Colector/genebank	Locality
Sapotaceae Juss.	Serrano (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 100 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 105 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 112 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 116 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 118 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 120 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 121 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 131 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 135 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 139 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 140 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 144 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 145 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 149 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 151 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 153 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 156 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 157 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 158 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 158 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 159 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 159 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 160 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 160 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 161 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 161 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 162 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 162 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 164 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 165 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 165 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 166 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 167 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 168 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 168 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 169 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 170 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 173 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 174 (UDBC)	Macarena

Taxon	Colector/genebank	Locality
Sapotaceae Juss.	Serrano 175 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 176 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 177 (UDBC)	Macarena
Sapotaceae Juss.	Serrano 18 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 185 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 193 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 198 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 203 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 204 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 205 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 207 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 208 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 209 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 210 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 211 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 212 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 215 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 218 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 219 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 219 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 221 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 223 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 224 (UDBC)	Catatumbo
Sapotaceae Juss.	Serrano 24 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 241 (UDBC)	Choco
Sapotaceae Juss.	Serrano 242 (UDBC)	Choco
Sapotaceae Juss.	Serrano 243 (UDBC)	Choco
Sapotaceae Juss.	Serrano 244 (UDBC)	Choco
Sapotaceae Juss.	Serrano 245 (UDBC)	Choco
Sapotaceae Juss.	Serrano 246 (UDBC)	Choco
Sapotaceae Juss.	Serrano 247 (UDBC)	Choco
Sapotaceae Juss.	Serrano 249 (UDBC)	Choco
Sapotaceae Juss.	Serrano 258 (UDBC)	Choco
Sapotaceae Juss.	Serrano 259 (UDBC)	Choco
Sapotaceae Juss.	Serrano 260 (UDBC)	Choco
Sapotaceae Juss.	Serrano 263 (UDBC)	Choco
Sapotaceae Juss.	Serrano 267 (UDBC)	Choco
Sapotaceae Juss.	Serrano 268 (UDBC)	Choco
Sapotaceae Juss.	Serrano 269 (UDBC)	Choco
Sapotaceae Juss.	Serrano 278 (UDBC)	Choco

Taxon	Colector/genebank	Locality
Sapotaceae Juss.	Serrano 28 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 280 (UDBC)	Choco
Sapotaceae Juss.	Serrano 284 (UDBC)	Choco
Sapotaceae Juss.	Serrano 285 (UDBC)	Choco
Sapotaceae Juss.	Serrano 286 (UDBC)	Choco
Sapotaceae Juss.	Serrano 288 (UDBC)	Choco
Sapotaceae Juss.	Serrano 294 (UDBC)	Choco
Sapotaceae Juss.	Serrano 296 (UDBC)	Choco
Sapotaceae Juss.	Serrano 298 (UDBC)	Choco
Sapotaceae Juss.	Serrano 299 (UDBC)	Choco
Sapotaceae Juss.	Serrano 301 (UDBC)	Choco
Sapotaceae Juss.	Serrano 303 (UDBC)	Choco
Sapotaceae Juss.	Serrano 305 (UDBC)	Choco
Sapotaceae Juss.	Serrano 306 (UDBC)	Choco
Sapotaceae Juss.	Serrano 307 (UDBC)	Choco
Sapotaceae Juss.	Serrano 309 (UDBC)	Choco
Sapotaceae Juss.	Serrano 311 (UDBC)	Choco
Sapotaceae Juss.	Serrano 312 (UDBC)	Choco
Sapotaceae Juss.	Serrano 315 (UDBC)	Choco
Sapotaceae Juss.	Serrano 318 (UDBC)	Choco
Sapotaceae Juss.	Serrano 319 (UDBC)	Choco
Sapotaceae Juss.	Serrano 320 (UDBC)	Choco
Sapotaceae Juss.	Serrano 321 (UDBC)	Choco
Sapotaceae Juss.	Serrano 334 (UDBC)	Choco
Sapotaceae Juss.	Serrano 335 (UDBC)	Choco
Sapotaceae Juss.	Serrano 336 (UDBC)	Choco
Sapotaceae Juss.	Serrano 337 (UDBC)	Choco
Sapotaceae Juss.	Serrano 338 (UDBC)	Choco
Sapotaceae Juss.	Serrano 339 (UDBC)	Choco
Sapotaceae Juss.	Serrano 348 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 350 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 351 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 352 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 355 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 357 (UDBC)	Magdalena Valley

Taxon	Colector/genebank	Locality
Sapotaceae Juss.	Serrano 359 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 38 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 39 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 44 (UDBC)	Magdalena Valley
Sapotaceae Juss.	Serrano 49 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 50 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 52 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 53 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 54 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 55 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 56 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 58 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 59 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 62 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 64 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 66 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 67 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 68 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 69 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 70 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 71 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 72 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 73 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 76 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 77 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 78 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 89 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 90 (UDBC)	Amazon
Sapotaceae Juss.	Serrano 98 (UDBC)	Amazon
Sapotaceae Juss.	TBC	Guyanas
Sarcaulus brasiliensis (A. DC.) Eyma	Paniagua *et al*. 4852 (MO)	Amazon
Sarcaulus brasiliensis (A. DC.) Eyma	Richardson *et al*. 322 (E)	Amazon
Sarcosperma laurinum (Benth.) Hook. f.	Saunders s.n. (S) AM408055	Asia
Synsepalum dulcificum (Schumach. & Thonn.) Daniell	Welsing *et al*. 24 (WAG) DQ246697	Africa

Taxon	Colector/genebank	Locality
Synsepalum passargei (Engl.) T.D. Penn.	Magogo 2452 (UPS) DQ246698	Africa
Xantolis siamensis (Fletcher) P. Royen	Smitairi 1 (L)	Asia

**TBC: to be confirmed*

Appendix 2.1. **List of taxa included in the phylogenetic analyses**.
A total of 146 Chrysophylloideae accessions were collected in the Choco, Magdalena valley, Catatumbo, Amazon and Macarena lowland rain forests of Colombia. The remaining sequences were obtained from previous phylogenetic studies on Chrysophylloideae (Sánchez-C. *et al*., 2017; Gonzalez *et al*., 2009; Swenson *et al*., 2008) and GenBank.

Chapter 3. Linking Applied Conservation to Analyses of Species Distribution and Diversity Patterns across the Neotropical Lowland Rain Forest

3.1 Introduction

The lowland rain forest biome occurs below 1,000 m elevation and is generally characterised by evergreen vegetation adapted to relatively high and stable mean annual temperature and high annual precipitation (>2,000 mm per year) (Olson *et al.*, 2001; Burnham and Johnson, 2004; IDEAM *et al.,* 2007; Reynel *et al.,* 2013; Bernal *et al.,* 2016).

In the Neotropics this biome is also one of the richest, hosting 6,727 tree species in areas such as the Amazon (Cardoso *et al.,* 2017). This high biotic diversity, set against increasing threats of habitat degradation due to expanding human activity and land use change, highlights the conservation value of many lowland rain forest areas. Moreover, it emphasises the need for baseline biodiversity studies to identify conservation priority areas effectively, and to update current systems of protected areas at national scales.

One of the first and most important steps in any effective conservation planning is the identification of areas with distinctive biotas. Such work requires understanding of species distribution patterns across space and the ability to identify areas of endemism. A comprehensive effort aiming to explore distribution patterns within the Neotropics, including lowland rain forests, was that of Morrone (2001). Morrone's (2001) system, based on cladistics and parsimony analyses, divided the neotropical region into provinces nested within subregions. This approach allowed him to determine the supposed effects of vicariance on the distribution patterns of taxa, and to discover areas that may share a common biogeographic history. However, historical biogeographic studies like those of Morrone (2001) are based on evolutionary assumptions and ignore patterns of distribution caused by dispersal, contemporary environmental factors and by ecological interactions. They can consequently depict

only a partial picture about distribution patterns within areas. For instance, recent analyses (see Chapter 1) of biotic homogeneity within Morrone's (2001) units have demonstrated that this system does not accurately represent patterns of aggregation in a dominant group in the neotropical lowland rain forests, the Sapotaceae. It seems unlikely that Morrone's (2001) system would be a good model for conservation planning in other plant groups within this biome. The same could apply to other widely used biogeographic regionalisations which have not incorporated rigorous analyses of current patterns of distribution and diversity such as the terrestrial ecoregions (Olson *et al.,* 2001). The terrestrial ecoregions are not based on quantitative data, and were produced as approximations of distinct biotic units based on expert knowledge. These units are only available at low spatial resolution and are not applicable to national or regional level conservation planning (see Sarkinen *et al.,* 2011).

Other available systems such as the biodiversity hotspots (Myers *et al.,* 2000) have identified areas of high species richness and endemism under threat of habitat loss, with only 30% or less of their original extension remaining. The biodiversity hotspots provided an approximation of the status of areas considered of high current conservation value, yet for the Neotropics they appear too broad. Evidence for this has been found for example in studies based on floristic, spatial, and phylogenetic analyses (Bernal *et al.,*, 2016; Chapter 1 and Chapter 2) which suggest that there may be distinctive subunits within the large hotspots (787,760 km^2 of size in average), in areas such as the inter-Andean valleys in Colombia.

Although hotspots and biogeographic units are part of different initiatives, they both aim to create a framework that could serve to focus scientific research and conservation, and to create an implicit collaboration that ensures that local, national, and international resources are invested in the achievement of similar goals. This has attracted financial support and scientific interest that has positively reinforced the status of "high conservation priority" in several areas around the world. However, due to the size of the study units and the methodology used in these approximations, they are relevant mainly at global and continental scales. To achieve effective conservation planning, these larger biogeographic units must be complemented by research focused on local and national scales, and by analyses including current patterns of species

distribution. In many cases an appropriate scale is national, because this is a common political level for investment and conservation planning.

3.1.1 Zooming into national conservation priorities in the lowland rainforest biome: the Colombian case

Colombia is one of the richest countries in terms of species and ecosystem diversity. The origins of this diversity have been influenced by dispersal among areas of forest, and by tectonic activity resulting in the Andean uplift and the consequent creation of new habitats. The rise of the Andes isolated some populations on either side of its mountains, possibly driving biotic distinction among areas of lowland rain forest such as the Amazon, the Choco, Catatumbo, Macarena and the inter-Andean valleys (Chapter 1 and 2). These forests are some of the most biodiverse places on Earth, hosting about 50% of the over 22,000 species of vascular plants found in Colombia (Forero, 1988; FEN, 1993; Bernal *et al.,* 2016).

The Biogeographic Choco includes the lowland belt located between the Pacific Ocean and the Andean mountains in Colombia, Peru and Ecuador. In Colombia, where the Andes divide into three ranges, the Biogeographic Choco is found between the Western Cordillera and the Pacific coast. The landscapes there range from estuaries to flooded plains and hills below 1,200 m altitude, and ecosystems host almost the same number of species as the Colombian Amazon in an area six times smaller (Bernal *et al*, 2016). The biogeographic Choco is considered to be of high conservation value, it is represented in the Colombian system of protected areas by six national parks and one flora and fauna sanctuary, and is also included as one of the earth's hotspots proposed by Myers *et al.,* (2000).

The Catatumbo forests, geographically isolated from other lowland areas by the Eastern Cordillera and the Merida Andes, are some of the best preserved but least known forests in the Neotropics. Large stands of pristine rain forests can be found in this area with emergent layers that can reach up to 40 m (Chavez and Arango, 1997). Currently the Catatumbo forests are protected under Colombian and Venezuelan legislation as national parks. Also part of the Colombian system of national parks, La Macarena is the western-most highland of the Guyana shield, and it is located between

the eastern flank of the Eastern Cordillera of the Andes and the Amazon. This location suggests that La Macarena is a biotic island with high endemism and a unique assemblage which contains elements from the Andes, the Guianas and the Amazon (BirdLife International, 2017).

The inter-Andean valleys are highly degraded areas where selective logging and over-harvesting have caused the loss of most of the original extension of primary lowland rain forest. Nonetheless, the Magdalena valley has the third highest number of endemic plant species in Colombia, with 395 species known to be restricted to its forests (Bernal *et al.,* 2016).

The Amazon is the largest area of lowland rain forest in the Neotropics and has been studied extensively (e.g. Rangel-Ch, 2008; Londoño, 1997; ter Steege *et al.,* 2013; Cardoso *et al.,* 2017). It is one of the most biodiverse areas in the world, and in Colombia is home to 4,912 species of vascular plants (Bernal *et al.,* 2016). Eleven national parks within the Colombian system of protected areas are located in the Amazon.

Little is known about the origins, evolution and current distribution patterns of biotas within these distinct regions of lowland rain forest areas, and Colombian studies include only preliminary biogeographic systems and ecosystem classifications. Hernandez *et al.* (1992) developed a classification system for Colombian biomes using climatic, elevation and edaphic criteria, and incorporating the physiognomy and composition of undisturbed Colombian vegetation. This system served as an approximation for the delimitation of biogeographic units in the country, and was later used by the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM) to propose a new classification for Colombian ecosystems (IDEAM *et al.,* 2007). Ecosystems were drawn by intercepting Hernandez *et al.*'s (1992) biomes and land cover polygons from satellite imagery.

These approaches created a baseline for conservation initiatives, including the work of Corzo (2008), who proposed an update on biogeographic units and conservation priority areas focused on Colombian ecosystems (Fig. 3.1). In Corzo's (2008) system, Colombia was divided into biogeographic Districts based mainly on

the biogeographic units proposed by Hernandez *et al.,* (1992) and the ecosystems proposed by IDEAM *et al*., (2007). Areas of conservation priority were defined by assigning a value for representativeness, urgency, status and opportunity to each District. Representativeness was defined as how well the ecosystems within each District were represented by the Colombian system of protected areas, and how much was still needed to achieve the conservation goals established in the Convention on Biological Diversity (CBD) in 2004. Urgency was defined based on threats, vulnerability, and risk. Status was defined as natural or semi-natural, and opportunity was defined based on the availability of conservation studies or management plans.

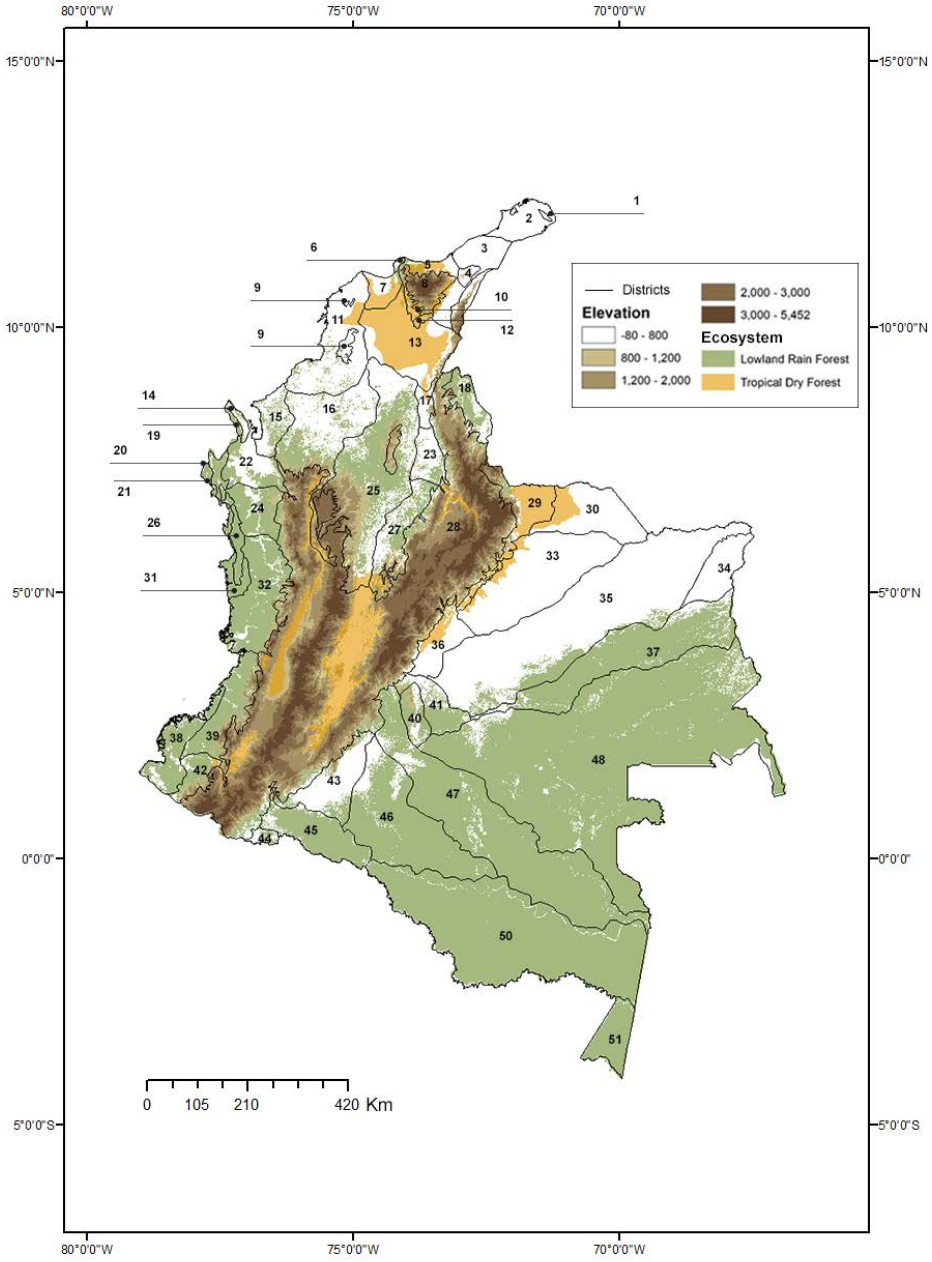


Figure 3.1. Corzo's (2008) Districts and the lowland rain forests of Colombia.
Black lines and numbers represent Corzo's (2008) Districts. 1: Macuira, 2: Alta Guajira, 3: Baja Guajira y Alto Cesar, 4: Marocaso, 5: Guachaca, 6: Santa Marta Enclaves Azonales, 7: Delta del Magdalena , 8: Chundua, 9: Maria y Piojo, 10: Aracataca, 11: Cartagena, 12: Caracolicito, 13: Ariguani-Cesar, 14: Acandi-San Blas, 15: Turbo, 16: Sinu-San Jorge, 17: La Gloria, 18: Catatumbo, 19: Tacarcuna, 20: Aspave-El Limon-Pirre, 21: Jurado, 22: Rio Sucio, 23: Lebrija, 24: Murri, 25: Nechi, 26: Baudo, 27: Carare, 28: Perija, 29: Arauca-Apure, 30: Piedemonte Casanare-Arauca, 31: Utria, 32: Alto Atrato-San Juan, 33: Casanare, 34: Maipures, 35: Sabanas Altas, 36: Piedemonte Meta, 37: Selvas del Norte de Guaviare, 38: Tumaco, 39: Micay, 40: Macarena, 41: Ariari-Guayabero, 42: Barbacoas, 43: Florencia, 44: Kofan, 45: Alto Putumayo, 46: Caguan, 47: Yari-Mariti, 48: Complejo Vaupes, 50: Huitoto, 51: Ticuna. Lowland rain forests are depicted as proposed by IDEAM *et al*., (2007), and dry forests areas are adapted from the WWF Ecorregions (Olson *et al*., 2001).

After evaluating each District according to those criteria, each unit was assigned to one of the categories shown in Table 3.1, in which the highest priority was given to areas absent from the national system of protected areas, facing a high level of threat, of natural status, owned by the Colombian government, and with available conservation plans. These areas and their categorisation are currently being used by governmental institutes leading conservation planning in Colombia (e.g. IDEAM, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Parques Nacionales), and by private consultancy companies who use them to characterise areas of economic interest prior to the exploitation of resources such as hydrocarbons, gold and coal (e.g. environmental impact assessments).

Table 3.1. Criteria used by Corzo (2008) in the definition of areas of priority for conservation. In Corzo's (2008) system Colombia was divided into biogeographic Districts, each District with a corresponding value for representativeness, urgency, status and opportunity (see text). According to those criteria, eight (VIII following Corzo, 2008) categories for conservation were defined, number I representing areas of main conservation concern, and number VIII representing areas of least conservation concern.

Category of conservation	Representativeness			Urgency		Status		Opportunity		
	Omission	insufficient	Sufficient	Urgent	non-urgent	Natural	Semi-natural	With existing conservation plans	Land owned by the state without existing conservation plans	No opportunity for the establishment of protected areas
I	x			x		x		x	x	
II	x			x		x				x
III	x			x			x	x	x	x
IV	x				x	x		x	x	x
V		x		x		x		x	x	x
VI		x			x	x		x	x	x
VII			x	x		x		x	x	x
VIII			x		x	x		x	x	x

The categorisation of conservation priority areas in Colombia has enhanced the understanding of local biotas, but similar to the issues identified in the biogeographic regionalisation of Morrone (2001; Chapter 1), the Colombian conservation priority categorisation plans have failed to consider current patterns of species distribution and richness. At the moment, despite efforts in Colombia to define preliminary priority areas for conservation, and although major investment has been made over the past decades through the establishment of large national protected areas (e.g., Chiribiquete National Park), a data-driven biogeographic division or ecoregionalisation of the lowland rain forests that would reflect distribution patterns and endemism in plant communities is still lacking. Incorporating such studies in conservation planning will guarantee that conservation units accurately represent patterns of species diversity and distribution within biomes.

Considering the high risk of biodiversity loss and the preliminary stages of conservation studies in countries like Colombia, the need to identify areas that could serve as basic units for conservation is urgent. We consider that an integrative approach that reflects species diversity patterns, endemism and phylogenetic diversity could be the best method to fill this gap. Hence, to complement the test of biogeographic units and the phylogenetic analyses carried out in chapters one and two, here we aim to take a step forward towards a new system for the delineation of areas of high conservation priority in the Neotropics by using the national level as a starting point. We use Species Distribution Models (SDMs) to explore patterns of habitat suitability for tree species in Sapotaceae, a key element in terms of species diversity and abundance across neotropical lowland rain forests, to identify areas of high predicted species richness in lowland rain forests across Colombia. Areas of high predicted species richness are compared to the current systems of national protected areas and to the conservation priority areas proposed by Corzo (2008).

3.2 Methods

We modelled the distribution of 202 neotropical Sapotaceae species, using a recently curated data set with 22,917 occurrence records obtained from the Herbario Nacional, Herbario Forestal, Herbario Amazónico Colombiano, Herbario de la

Universidad del Valle, Herbario "Choco" in Colombia, and the PADME, GBIF (Global Biodiveristy Information Facility) and TROPICOS® databases. This was performed in Maxent v.3.4.1 (Phillips *et al*., 2006, Phillips and Dudík, 2008). Maxent was used as it has been found to perform better than other algorithms and it allows the addition of techniques that correct for collection bias (Elith et al., 2006; Kramer-Schadt *et al.,* 2013). Maxent initially assumes the distribution of each species is the same as the distribution of environmental values, then iteratively solves for a constraining function for the distribution of the species given the discrepancy between the probability of environmental values in the background and the probability of environmental values in the species' presence sites (Elith *et al*., 2011). We assumed a cut off of five records per species, and species with fewer records were not used for modelling.

To control for unequal intensity in collection efforts, two sets of models were run, one using spatial filtering, and one using spatial filtering and calculating the Kernel Density Estimator (Wiegand and Moloney, 2014). The Kernel Density Estimator was calculated using a grid with a bandwidth bigger than that used in the modelling to smooth the density of occurrence records across the Neotropics. For this, the raster grid was generated using all the occurrences of vascular plants, except those of Sapotaceae, recorded in the Global Biodiversity Information Facility (GBIF). Spatial filtering was performed by decreasing the number of occurrences in oversampled areas (Kramer-Schadt *et al.,* 2013) using a radius of 10 km around each point. All analyses were performed in R 3.3.2.

Climatic predictors at a resolution of 2.5 arc-minutes were obtained from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) (Funk *et al.,* 2015) and the NASA Moderate Resolution Imagining Spectroradiometer land surface Temperature (MODIS v41) (Wan and Dozier, 1996; Wan, 2014) datasets. Expert knowledge, Principal Component Analyses, and regression analyses were used for variable selection to ensure the most informative and least correlated predictors were included with ≤0.7 Pearson correlation co-efficient threshold (Table 3.2). Among the 21 available variables, the following nine climatic predictors were used: isothermality, temperature seasonality, maximum temperature of warmest month, minimum

temperature of coldest month, temperature annual range, precipitation of driest month, precipitation seasonality, precipitation of wettest quarter and altitude.

Table 3.2. Climatic preditors used to model Sapotaceae especies in the Neotropics.
Climatic predictors at a resolution of 2.5 arc-minutes were downloaded from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) (Funk *et al*., 2015) and the NASA Moderate Resolution Imagining Spectroradiometer land surface Temperature (MODIS). The Pearson correlation co-efficient was calculated among pairs of predictors. In cases of high correlation (>0.7), the most biologically representative variable was chosen. (values <0.7 shown in bold). Alt. = Altitude, Het. = Heterogenity.

	Alt.	BIO1	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19	BIO2	BIO3	BIO4	BIO5	BIO6	BIO7	BIO8	BIO9	Het.
Alt.	1.00	-0.78	-0.66	-0.75	-0.19	-0.16	-0.17	0.20	-0.18	-0.18	-0.12	-0.20	0.57	0.46	0.13	-0.22	-0.78	0.42	-0.73	-0.70	**0.73**
BIO1	-0.78	1.00	**0.93**	**0.92**	-0.08	-0.05	-0.12	0.10	-0.03	-0.12	-0.15	0.02	-0.25	-0.38	-0.02	0.65	**0.81**	-0.12	**0.91**	**0.90**	-0.68
BIO10	-0.66	**0.93**	1.00	**0.74**	-0.25	-0.19	-0.26	0.22	-0.18	-0.27	-0.29	-0.14	0.00	-0.46	0.33	**0.84**	0.59	0.19	**0.87**	**0.82**	-0.62
BIO11	-0.75	**0.92**	**0.74**	1.00	0.08	0.11	-0.02	0.04	0.14	-0.02	-0.06	0.16	-0.46	-0.25	-0.39	0.39	**0.93**	-0.41	**0.79**	**0.89**	-0.61
BIO12	-0.19	-0.08	-0.25	0.08	1.00	**0.89**	**0.81**	-0.51	**0.92**	**0.84**	**0.79**	**0.83**	-0.41	0.09	-0.43	-0.44	0.17	-0.46	-0.18	0.03	-0.07
BIO13	-0.16	-0.05	-0.19	0.11	**0.89**	1.00	0.52	-0.14	**0.98**	0.57	0.58	**0.81**	-0.36	0.04	-0.40	-0.34	0.18	-0.39	-0.17	0.09	0.00
BIO14	-0.17	-0.12	-0.26	-0.02	**0.81**	0.52	1.00	-0.74	0.58	**0.99**	**0.80**	0.62	-0.33	0.09	-0.30	-0.43	0.07	-0.38	-0.17	-0.06	-0.13
BIO15	0.20	0.10	0.22	0.04	-0.51	-0.14	-0.74	1.00	-0.21	-0.75	-0.60	-0.30	0.39	0.00	0.22	0.44	-0.10	0.42	0.08	0.11	0.18
BIO16	-0.18	-0.03	-0.18	0.14	**0.92**	**0.98**	0.58	-0.21	1.00	0.62	0.62	**0.84**	-0.38	0.04	-0.43	-0.35	0.21	-0.42	-0.16	0.11	-0.05
BIO17	-0.18	-0.12	-0.27	-0.02	**0.84**	0.57	0.99	-0.75	0.62	1.00	**0.83**	0.64	-0.35	0.10	-0.32	-0.45	0.08	-0.40	-0.17	-0.06	-0.12
BIO18	-0.12	-0.15	-0.29	-0.06	**0.79**	0.58	0.80	-0.60	0.62	**0.83**	1.00	0.50	-0.24	0.23	-0.29	-0.41	0.00	-0.31	-0.12	-0.17	-0.05
BIO19	-0.20	0.02	-0.14	0.16	**0.83**	**0.81**	0.62	-0.30	0.84	0.64	0.50	1.00	-0.40	-0.01	-0.39	-0.33	0.25	-0.44	-0.19	0.20	-0.08
BIO2	0.57	-0.25	0.00	-0.46	-0.41	-0.36	-0.33	0.39	-0.38	-0.35	-0.24	-0.40	1.00	0.38	0.64	0.52	-0.72	**0.94**	-0.16	-0.31	0.29
BIO3	0.46	-0.38	-0.46	-0.25	0.09	0.04	0.09	0.00	0.04	0.10	0.23	-0.01	0.38	1.00	-0.28	-0.22	-0.33	0.08	-0.33	-0.37	0.40
BIO4	0.13	-0.02	0.33	-0.39	-0.43	-0.40	-0.30	0.22	-0.43	-0.32	-0.29	-0.39	0.64	-0.28	1.00	0.59	-0.51	0.83	0.08	-0.13	0.01
BIO5	-0.22	0.65	**0.84**	0.39	-0.44	-0.34	-0.43	0.44	-0.35	-0.45	-0.41	-0.33	0.52	-0.22	0.59	1.00	0.13	0.66	0.63	0.53	-0.35
BIO6	-0.78	**0.81**	0.59	**0.93**	0.17	0.18	0.07	-0.10	0.21	0.08	0.00	0.25	-0.72	-0.33	-0.51	0.13	1.00	-0.66	0.67	**0.81**	-0.57
BIO7	0.42	-0.12	0.19	-0.41	-0.46	-0.39	-0.38	0.42	-0.42	-0.40	-0.31	-0.44	**0.94**	0.08	**0.83**	0.66	-0.66	1.00	-0.03	-0.21	0.17
BIO8	-0.73	**0.91**	**0.87**	**0.79**	-0.18	-0.17	-0.17	0.08	-0.16	-0.17	-0.12	-0.19	-0.16	-0.33	0.08	0.63	0.67	-0.03	1.00	**0.71**	-0.65
BIO9	-0.70	**0.90**	**0.82**	**0.89**	0.03	0.09	-0.06	0.11	0.11	-0.06	-0.17	0.20	-0.31	-0.37	-0.13	0.53	**0.81**	-0.21	**0.71**	1.00	-0.57
Het.	**0.73**	-0.68	-0.62	-0.61	-0.07	0.00	-0.13	0.18	-0.05	-0.12	-0.05	-0.08	0.29	0.40	0.01	-0.35	-0.57	0.17	-0.65	-0.57	1.00

The region used to train the models was delineated as the neotropical region (Morrone, 2001) to provide a biologically realistic modelling extent. Each species was modelled five times controlling for bias using spatial filtering plus the Kernel Density Estimator, and five times using the Kernel Density Estimator. Models were trained with 80% of the records for a species and tested using the remaining 20% of the sites. We used the continuous Boyce Index (CBI; Boyce *et al.,* 2002, Hirzel *et al.,* 2006), which is the rank correlation coefficient between the proportion of predictions across the study region within a given range with the proportion of predicted values at presence sites with predictions in the same range. We used overlapping ranges 0.1 in width to calculate proportions of sites (Hirzel *et al.,* 2006). The CBI was specifically developed for assessing the performance of models when test absences are unavailable. It reflects a model's ability to assign presence sites predictions higher than would be expected were the same set of predictions located randomly across the landscape. It ranges from -1 to 1, with positive values indicating better performance. Others have used area under the receiver operator curve (AUC) for evaluating presence-only models, but AUC has been shown to be an unreliable indicator of model performance (Li and Guo, 2013; Smith, 2013). CBI was calculated for each of the five iterations and then used to calculate the mean CBI value per species per set of models. One model set per species was chosen as the one with the highest mean CBI value among the two sets of models that were run (Table 3.3). The best models were then summed to produce an estimated species richness map for Sapotaceae in the Neotropics. This map was used to identify areas of high predicted species richness within Colombia. Predicted species richness per District and national protected areas across Colombia were measured by using two different diversity metrics: (1) the maximum number of species per pixel within each District *sensu* Corzo (2008) (2) the mean number of species per pixel within each District. We also measured the number of endemic species for each District, referring to the number of species that are restricted to a single District.

All figures in this chapter were produced using ArcMap 10.1 and Photoshop CS6.

Table 3.3. **Mean values for the Continuous Boyce Index (CBI) for species distribution models of neotropical Sapotaceae**.
Mean CBI values were calculated for model set 1 and 2. Model set 1 included each of the five iterations where bias in collection density was controlled using spatial filtering and the Kernel Density Estimator. Model set 2 included each of the five iterations where bias in collection density was controlled using only the Kernel Density Estimator. Mean CBI values for model set 1 and 2 were calculated and compared to identify which set of models performed better. The model set with highest mean CBI value (in bold) was used for each species. Species richness across Colombia was summarised by summing the thresholded best models for all species. Only models with CBI >0.5 were used in the analyses of species richness.

Species	CBI model set 1 (mean)	CBI model set 2 (mean)
Chrysophyllum amazonicum	**0.84**	0.65
Chrysophyllum argenteum	**0.99**	0.67
Chrysophyllum bombycinum	**0.52**	0.42
Chrysophyllum brenesii	**0.57**	0.56
Chrysophyllum caimito	**0.71**	0.57
Chrysophyllum cainito	**0.94**	0.39
Chrysophyllum colombianum	**0.68**	0.18
Chrysophyllum cuneifolium	**0.66**	0.34
Chrysophyllum flexuosum	**0.63**	0.58
Chrysophyllum gonocarpum	**1.00**	0.82
Chrysophyllum lucentifolium	**0.75**	0.59
Chrysophyllum manaosense	**0.67**	0.56
Chrysophyllum marginatum	**0.96**	0.80
Chrysophyllum mexicanum	**0.93**	0.79
Chrysophyllum oliviforme	0.78	**0.85**
Chrysophyllum pomiferum	0.66	**0.67**
Chrysophyllum prieurii	**0.83**	0.62
Chrysophyllum sanguinolentum	**0.76**	0.64
Chrysophyllum sparsiflorum	**0.81**	0.58
Chrysophyllum splendens	**0.83**	0.63
Chrysophyllum venezuelanense	**0.98**	0.73
Chrysophyllum viride	**0.54**	0.25
Diploon cuspidatum	**0.58**	0.12
Ecclinusa bullata	**0.57**	0.49
Ecclinusa guianensis	**0.73**	0.54
Ecclinusa lanceolata	**0.93**	0.67
Ecclinusa ramiflora	**0.79**	0.41
Elaeoluma glabrescens	**0.90**	0.73
Elaeoluma nuda	**0.66**	0.54
Elaeoluma schomburgkiana	**0.63**	0.50
Manilkara bidentata	**0.94**	0.71
Manilkara chicle	**0.91**	0.77
Manilkara huberi	**0.62**	0.57
Manilkara longifolia	**0.58**	0.46

Species	CBI model set 1 (mean)	CBI model set 2 (mean)
Manilkara salzmannii	**0.66**	0.62
Manilkara zapota	**0.93**	0.72
Micropholis crassipedicellata	0.55	**0.57**
Micropholis crotonoides	0.55	**0.58**
Micropholis egensis	**0.97**	0.84
Micropholis gardneriana	**0.80**	-0.09
Micropholis guyanensis	**0.97**	0.88
Micropholis madeirensis	0.52	**0.53**
Micropholis melinoniana	**0.80**	0.58
Micropholis porphyrocarpa	**0.69**	0.27
Micropholis truncifiora	0.53	**0.55**
Micropholis venulosa	**0.99**	0.93
Micropholis williamii	**0.58**	0.54
Mimusops coriacea	**0.62**	0.59
Pouteria ambelaniifolia	0.52	**0.65**
Pouteria baehniana	**0.76**	0.46
Pouteria bangii	**0.87**	0.37
Pouteria bilocularis	**0.97**	0.82
Pouteria bracteata	**0.56**	0.37
Pouteria buenaventurensis	0.61	**0.75**
Pouteria caimito	**0.96**	0.90
Pouteria campanulata	**0.64**	0.47
Pouteria campechiana	0.90	**0.92**
Pouteria cladantha	**0.87**	0.67
Pouteria cuspidata	**0.98**	0.82
Pouteria egregia	**0.71**	0.46
Pouteria elegans	**0.90**	0.82
Pouteria eugeniifolia	**0.72**	0.60
Pouteria franciscana	0.49	**0.50**
Pouteria gabrielensis	0.49	**0.53**
Pouteria gardneri	**0.84**	0.19
Pouteria gardneriana	**0.89**	0.31
Pouteria glomerata	**0.99**	0.74
Pouteria gomphiifolia	**0.83**	0.74
Pouteria grandiflora	**0.70**	0.59
Pouteria guianensis	**0.96**	0.89
Pouteria hispida	**0.79**	0.57
Pouteria izabalensis	0.20	**0.62**
Pouteria juruana	**0.53**	0.36
Pouteria krukovii	**0.74**	0.74
Pouteria laevigata	0.73	**0.78**
Pouteria leptopedicellata	**0.61**	0.58

Species	CBI model set 1 (mean)	CBI model set 2 (mean)
Pouteria lucuma	**0.73**	0.14
Pouteria macahensis	**0.51**	0.17
Pouteria macrocarpa	**0.50**	0.26
Pouteria macrophylla	**0.94**	0.54
Pouteria multiflora	**0.94**	0.83
Pouteria oblanceolata	**0.63**	0.33
Pouteria pachyphylla	0.48	**0.67**
Pouteria platyphylla	**0.59**	0.58
Pouteria procera	**0.64**	0.27
Pouteria ramiflora	**1.00**	0.63
Pouteria reticulata	**0.98**	0.74
Pouteria retinervis	**0.56**	0.47
Pouteria rostrata	**0.80**	0.74
Pouteria sagotiana	**0.53**	0.49
Pouteria sapota	**0.92**	0.70
Pouteria speciosa	0.40	**0.52**
Pouteria subcaerulea	**0.77**	-0.02
Pouteria subrotata	**0.64**	0.43
Pouteria surumuensis	0.45	**0.63**
Pouteria trigonosperma	**0.86**	0.46
Pouteria trilocularis	**0.74**	0.52
Pouteria ucuqui	**0.79**	0.66
Pouteria venosa	**0.77**	0.22
Pouteria vernicosa	0.29	**0.67**
Pouteria virescens	**0.70**	0.23
Pradosia beardii	**0.51**	0.40
Pradosia brevipes	**0.59**	-0.21
Pradosia grisebachii	**0.52**	0.28
Pradosia ptychandra	**0.80**	0.37
Pradosia schomburgkiana	0.61	**0.68**
Pradosia surinamensis	**0.93**	0.34
Sarcaulus brasiliensis	0.48	**0.55**
Sarcaulus inflexus	**0.72**	0.56
Sideroxylon capiri	0.50	**0.58**
Sideroxylon floribundum	**0.86**	0.29
Sideroxylon foetidissimum	0.27	**0.76**
Sideroxylon obovatum	**0.97**	0.27
Sideroxylon obtusifolium	0.41	**0.89**
Sideroxylon palmeri	**0.76**	0.32
Sideroxylon portoricense	**0.82**	0.51
Sideroxylon salicifolium	0.56	**0.68**
Sideroxylon stenospermum	0.52	**0.69**

3.3 Results

Corzo (2008) proposed areas of conservation priority in Colombia first by delineating biogeographic Districts and then by evaluating each District according to criteria such as representativeness. Here, Corzo's (2008) Districts were compared to areas of high predicted species richness in Sapotaceae to evaluate the accuracy by which Corzo's (2008) Districts depict patterns of distribution in the lowland rain forest of Colombia.

Occurrence data of Sapotaceae were gathered for a total of 225 species, after excluding species with less than five unique records, and after correcting for collection bias, 202 species were selected for modelling. Models were then evaluated using the continuous Boyce index (CBI), setting a threshold of 0.5. For 84 species, both model sets (see Methods) fell below the CBI threshold of 0.5 and these species were discarded from further analyses due to poor model performance. Model set 1 had the highest CBI (better performance) in 95 out of the final 118 species. In this model set, controlling for collection bias was performed by using spatial filtering and by calculating the Kernel Density Estimator. The mean CBI for model set 1 in all species was of 0.52, and the mean CBI for model set 2 in all species was of 0.4. The mean CBI for models with good performance (>0.5) in model set 1 was of 0.71, and the mean CBI for models with good performance (>0.5) in model set 2 was of 0.54. For the final 118 species (Table 3.3; Appendix 3.1), the model set with highest CBI value was thresholded using the 10 percentile training presence criterion. The thresholded models were then summed to produce a map of estimated species richness for Sapotaceae across Colombia within Corzo's (2008) Districts (Fig. 3.2). Areas of high diversity were compared to the current system of protected areas in Colombia.

Species richness within Corzo's (2008) units was measured using two different diversity measures, but we focus here on presenting results based on the maximum potential number of Sapotaceae species per pixel within each District. This is because the measure is standardized for area (i.e. pixel), and does not suffer from the artificial inflation of species diversity values that occurs when using the mean number of species per pixel within each District or when using the number of species per District in cases where district boundaries include various ecosystems. Examples of such issues include

the Sinu-San Jorge District, where the number of species per District is high when calculated using the polygon boundaries, but more moderate and comparable to other Districts when using the maximum number of species per pixel (Appendix 3.2). Endemism was also measured within each District, but because endemism in Sapotaceae was found to be zero across all Districts, we do not report these results in the main tables.

Focusing on results from the maximum number of species per pixel, our analyses show that the Selvas del Norte de Guaviare (37 in Fig. 3.2), and Complejo Vaupes (48 in Fig. 3.2) Districts in the Amazon, and the Nechi District (25 in Fig. 3.2) in the Magdalena inter-Andean valley (Fig. 3.2 and Table 3.4) are the most species rich with 113, 113 and 112 species respectively (Fig. 3.2 and Table 3.4). These Districts are followed by Sabanas Altas (35 in Fig. 3.2) that mainly covers the Colombian savannahs and the gallery forest of the Vichada River with 111 species (Fig. 3.2 and Table 3.4).

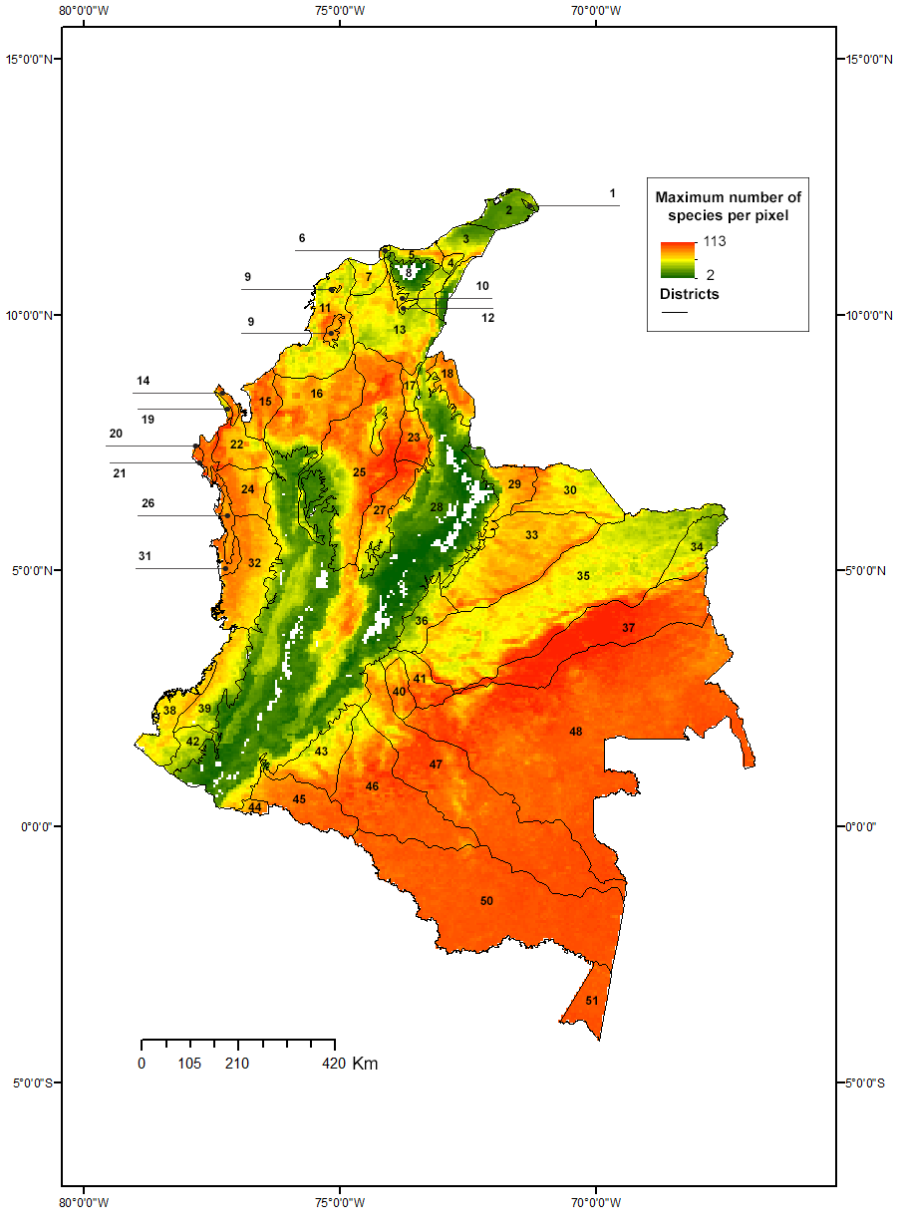


Figure 3.2. Predicted species richness within Corzo's (2008) units.
Predicted species richness of neotropical Sapotaceae was measured as the maximum number of Sapotaceae species per pixel within each District, and it is represented by colours from red to green. Red colours represent areas of high predicted species richness and green colours areas of low predicted species richness based on SDM results. Black lines and numbers represent Corzo's (2008) Districts. 1: Macuira, 2: Alta Guajira, 3: Baja Guajira y Alto Cesar, 4: Marocaso, 5: Guachaca, 6: Santa Marta Enclaves Azonales, 7: Delta del Magdalena , 8: Chundua, 9: Maria y Piojo, 10: Aracataca, 11: Cartagena, 12: Caracolicito, 13: Ariguani-Cesar, 14: Acandi-San Blas, 15: Turbo, 16: Sinu-San Jorge, 17: La Gloria, 18: Catatumbo, 19: Tacarcuna, 20: Aspave-El Limon-Pirre, 21: Jurado, 22: Rio Sucio, 23: Lebrija, 24: Murri, 25: Nechi, 26: Baudo, 27: Carare, 28: Perija, 29: Arauca-Apure, 30: Piedemonte Casanare-Arauca, 31: Utria, 32: Alto Atrato-San Juan, 33: Casanare, 34: Maipures, 35: Sabanas Altas, 36: Piedemonte Meta, 37: Selvas del Norte de Guaviare, 38: Tumaco, 39: Micay, 40: Macarena, 41: Ariari-Guayabero, 42: Barbacoas, 43: Florencia, 44: Kofan, 45: Alto Putumayo, 46: Caguan, 47: Yari-Mariti, 48: Complejo Vaupes, 50: Huitoto, 51: Ticuna.

Other areas of lowland rain forest such as the Choco, Catatumbo and Macarena, which have been identified as regions with distinct biotas in previous analyses (e.g. Chapter 1 and 2; Myers, 2000; BirdLife International, 2017) show lower maximum species richness per pixel compared to the Selvas del Norte del Guaviare, Complejo Vaupes and Nechi Districts. The biogeographic Choco is divided into 13 units according to Corzo (2008) (Fig. 3.2). Within these 13 Districts the maximum number of species per pixel is 110 in the Rio Sucio District (22 in Fig. 3.2), which is the fifth richest unit in Colombia. The Catatumbo area is represented in Corzo's (2008) system by a single unit (18 in Fig. 3.2), with a maximum of 98 species per pixel, and is as the 27th richest District in Colombia. The Macarena lowland rain forests are divided into three Districts in Corzo's (2008) system, of which the Yari-Mariti District (47 in Fig. 3.2) has the highest species richness with a maximum number of species per pixel of 109. Yari-Mariti is the 10th richest area in Sapotaceae species in Colombia.

Table 3.4. Species richness within Corzo's (2008) Districts.
Species richness was measured as the maximum number of potential Sapotaceae species per pixel within each District by summarising species distribution models of Sapotaceae and intercepting that summary with the biogeographic units defined by Corzo (2008).

ID (Fig. 2)	District	Region of Colombia	Maximum number of species per pixel within each District
1	Macuira	Caribbean	36
2	Alta Guajira	Caribbean	46
3	Baja Guajira y Alto Cesar	Caribbean	85
4	Marocaso	Caribbean	80
5	Guachaca	Caribbean	86
6	Santa Marta Enclaves Azonales	Caribbean	77
7	Delta del Magdalena	Caribbean	75
8	Chundua	Caribbean	63
9	Maria y Piojo	Caribbean	93
10	Aracataca	Caribbean	72
11	Cartagena	Caribbean	102

ID (Fig. 2)	District	Region of Colombia	Maximum number of species per pixel within each District
12	Caracolicito	Caribbean	70
13	Ariguani-Cesar	Caribbean	102
14	Acandi-San Blas	Choco	103
15	Turbo	Choco	104
16	Sinu-San Jorge	Choco/Caribbean	109
17	La Gloria	Caribbean	97
18	Catatumbo	Catatumbo	98
19	Tacarcuna	Choco	87
20	Aspave-El Limon-Pirre	Choco	107
21	Jurado	Choco	107
22	Rio Sucio	Choco	110
23	Lebrija	Magdalena valley	110
24	Murri	Choco	101
25	Nechi	Magdalena valley/Caribbean	112
26	Baudo	Choco	95
27	Carare	Magdalena valley	110
28	Perija	Andes	96
29	Arauca-Apure	Llanos	101
30	Piedemonte Casanare-Arauca	Llanos	101
31	Utria	Choco	96
32	Alto Atrato-San Juan	Choco	91
33	Casanare	Llanos	101
34	Maipures	Llanos	96
35	Sabanas Altas	Llanos	111
36	Piedemonte Meta	Llanos	91
37	Selvas del Norte de Guaviare	Amazon	113
38	Tumaco	Choco	80
39	Micay	Choco	80
40	Macarena	Macarena	105
41	Ariari-Guayabero	Macarena	108
42	Barbacoas	Choco	62
43	Florencia	Amazon	100
44	Kofan	Amazon	89

ID (Fig. 2)	District	Region of Colombia	Maximum number of species per pixel within each District
45	Alto Putumayo	Amazon	103
46	Caguan	Amazon	110
47	Yari-Mariti	Amazon/Macarena	109
48	Complejo Vaupes	Amazon	113
50	Huitoto	Amazon	103
51	Ticuna	Amazon	105

In Colombia there are 470 protected areas that cover 164,617 km^2, equivalent to 14.4% of the total area of the country (IUCN, 2016; Table 3.5). Within the Districts that are most species-rich for Sapotaceae, one forest reserve has been established in the Selvas del Norte de Guaviare District covering a total of 0.76 km^2 equivalent to <1% of the total area of this District (Fig. 3.3, Table 3.5). In the Complejo Vaupes District, six protected areas cover 26,196 km^2, equivalent to 15.6% of the entire District. In the Nechi District, 18 protected areas have been declared (Fig. 3.3, Table 3.5). They include one national park, one regional park, one recreational area, seven integrated management districts, seven regional protected forest reserves, and one natural reserve of the civil society (Fig. 3.3, Table 3.5).

Table 3.5. Protected areas in Colombia (IUCN, 2016).
Eighteen types of protected areas have been established in Colombia. Natural Reserves of the Civil Society are privately owned and managed, and Indigenous Areas are owned and managed by indigenous tribes. Ecological Reserves, Forest Reserves, UNESCO-MAB Biosphere Reserves and World Heritage Sites have not officially reported management.

Districts	Area (Km^2)	Coverage of Protected Areas per District (%)
Alta Guajira	**90.37**	**1.22**
Integrated Management Regional District	**0.04**	**0.00**
Musichi	0.04	0.00
Natural National Park	**90.32**	**1.22**
Bahia Portete Kaurrele	55.17	0.74
Macuira	35.15	0.47
Alto Atrato-San Juan	**469.77**	**2.17**
National Protected Forest Reserve	**21.60**	**0.10**
Rio Dagua	21.60	0.10
Integrated Management Regional District	**68.16**	**0.31**
Territorio Colectivo del Consejo Comunitario de la Plata	68.16	0.31
Natural National Park	**130.50**	**0.60**
Tatama	68.79	0.32
Uramba Bahia Malaga	61.70	0.28
Natural Regional Park	**249.51**	**1.15**
la Sierpe	249.48	1.15
Paramo del Duende	0.03	0.00
Alto Putumayo	**4,526.33**	**29.01**
Natural National Park	**4,526.33**	**29.01**
la Paya	4,424.10	28.35
Serrania de Los Churumbelos	102.23	0.66
Aracataca	**13.05**	**0.85**
Natural National Park	**7.07**	**0.46**
Sierra Nevada de Santa Marta	7.07	0.46
Natural Reserve of the Civil Society	**5.97**	**0.39**
las Aves El Dorado	5.94	0.39
Pachamama	0.03	0.00
Ariari-Guayabero	**3,200.77**	**38.70**
Natural National Park	**3,200.77**	**38.70**
Sierra de la Macarena	3,200.77	38.70
Ariguani-Cesar	**606.46**	**2.56**
Integrated Management Regional District	**558.01**	**2.36**

Complejo Cienagoso de Zarate Malibu y Veladero	558.01	2.36
Natural National Park	**48.46**	**0.20**
Catatumbo - Bari	48.46	0.20
Baja Guajira y Alto Cesar	**152.31**	**1.48**
Fauna and Flora Sanctuary	**45.35**	**0.44**
Los Flamencos	45.35	0.44
Regional Protected Forest Reserve	**59.61**	**0.58**
Montes de Oca	59.61	0.58
Integrated Management Regional District	**14.87**	**0.14**
Musichi	14.87	0.14
Natural Regional Park	**28.67**	**0.28**
Los Besotes	28.67	0.28
Natural Reserve of the Civil Society	**3.81**	**0.04**
Paraver	3.81	0.04
Barbacoas	**27.56**	**0.49**
National Protected Forest Reserve	**27.56**	**0.49**
la Planada	2.75	0.05
Rio Nembi	24.81	0.45
Baudo	**453.10**	**15.56**
Natural National Park	**453.10**	**15.56**
Utria	453.10	15.56
Caguan	**6,468.61**	**16.09**
Natural National Park	**6,468.61**	**16.09**
Serrania de Chiribiquete	6,468.61	16.09
Carare	**2,367.52**	**17.65**
National Protected Forest Reserve	**130.25**	**0.97**
Cuchilla El Minero	101.79	0.76
Cuenca Hidrografica del Rio San Francisco	28.45	0.21
Integrated Management Regional District	**1,866.91**	**13.91**
Cuchilla de San Antonio	131.87	0.98
Rio Minero	350.94	2.62
Humedal San Silvestre	687.23	5.12
Serrania de Los yariguies	694.96	5.18
laguna del Coco	1.90	0.01
Natural Regional Park	**355.98**	**2.65**
Serrania de las Quinchas	355.98	2.65
Natural Reserve of the Civil Society	**14.38**	**0.11**
El Paujil	14.36	0.11
Ecosistemas Andinos	0.02	0.00
Casanare	**221.75**	**0.58**

Natural Reserve of the Civil Society	**221.75**	**0.58**
Hato Venecia de Guanapalo	65.50	0.17
la Aurora	99.69	0.26
la Esmeralda	19.18	0.05
las Malvinas	6.48	0.02
Matesanto	8.10	0.02
Palmarito	22.81	0.06
Catatumbo	**876.68**	**12.83**
Natural National Park	**876.68**	**12.83**
Catatumbo - Bari	876.68	12.83
Chundua	**3,250.49**	**53.46**
Natural National Park	**3,247.45**	**53.41**
Sierra Nevada de Santa Marta	3,247.45	53.41
Natural Regional Park	**3.05**	**0.05**
Los Besotes	3.05	0.05
Complejo Vaupes	**26,196.61**	**15.58**
Indigenous Area	2,087.63	1.24
Alto Rio Negro	2,087.63	1.24
Natural National Park	**4,205.73**	**2.50**
Sierra de la Macarena	18.69	0.01
yaigoje Apaporis	4,187.04	2.49
Natural Reserve	**19,902.48**	**11.84**
Nukak	8,910.80	5.30
Puinawai	10,991.68	6.54
Recreational Area	**0.77**	**0.00**
Cuatro Microcuencas	0.77	0.00
Acandi-San Blas	**338.39**	**25.49**
National Protected Forest Reserve	**211.48**	**15.93**
Darien	211.48	15.93
Natural National Park	**63.49**	**4.78**
Los Katios	63.49	4.78
World Heritage Site	**63.42**	**4.78**
Los Katios	63.42	4.78
Cartagena	**1,336.59**	**5.95**
Fauna and Flora Sanctuary	**17.76**	**0.08**
El Corchal "el Mono Hernandez"	15.73	0.07
Los Colorados	2.03	0.01
National Protected Forest Reserve	**62.39**	**0.28**
Serrania de la Coraza y Montes de Maria	62.39	0.28
Integrated Management Regional District	**1,188.56**	**5.29**
Complejo Cienagoso de Zarate Malibu y Veladero	87.27	0.39

Complejo Cenagoso del Bajo Sinu	802.88	3.58
Luriza	8.54	0.04
Sabanas Abiertas y Arbustivas-Galeras	16.58	0.07
Manglar de la Bahia de Cispata	255.17	1.14
Cienaga de la Caimanera	18.11	0.08
Natural National Park	**16.24**	**0.07**
Corales del Rosario y de San Bernardo	16.24	0.07
Natural Regional Park	**43.30**	**0.19**
Boca de Guacamaya	29.00	0.13
Los Rosales	14.30	0.06
Park Way	**1.69**	**0.01**
Isla de Salamanca	1.69	0.01
Natural Reserve of the Civil Society	**6.67**	**0.03**
Sanguare	6.67	0.03
Florencia	**110.52**	**0.73**
Natural National Park	**110.52**	**0.73**
Serrania de Los Churumbelos	110.52	0.73
Maria y Piojo	**13.88**	**0.87**
Fauna and Flora Sanctuary	**8.64**	**0.54**
Los Colorados	8.64	0.54
National Protected Forest Reserve	**5.24**	**0.33**
Serrania de la Coraza y Montes de Maria	5.24	0.33
Santa Marta Enclaves Azonales	**119.60**	**28.76**
Natural National Park	**118.86**	**28.58**
Tayrona	118.86	28.58
Natural Reserve of the Civil Society	**0.74**	**0.18**
La Iguana Verde	0.60	0.15
Parque Ambiental Palangana	0.14	0.03
Piedemonte Casanare-Arauca	**1,030.35**	**7.83**
National Protected Forest Reserve	**51.99**	**0.39**
Cuenca Alta del Rio Satoca	42.00	0.32
Quebrada la Tablona	4.35	0.03
Rio Tame	5.64	0.04
Natural National Park	**948.96**	**7.21**
El Cocuy	447.92	3.40
El Tama (Venezuela)	501.04	3.81
Natural Regional Park	**29.40**	**0.22**
Tablona	3.96	0.03
San Miguel de Los Farallones	25.44	0.19
Delta del Magdalena	**634.30**	**14.01**
Fauna and Flora Sanctuary	**282.10**	**6.23**

Cienaga Grande de Santa Marta	282.10	6.23
Park Way	**352.20**	**7.78**
Isla de Salamanca	352.20	7.78
Guachaca	**960.39**	**29.23**
Fauna and Flora Sanctuary	**31.39**	**0.96**
Los Flamencos	31.39	0.96
National Protected Forest Reserve	**2.99**	**0.09**
Jirocasaca	2.99	0.09
Natural National Park	**925.93**	**28.18**
Sierra Nevada de Santa Marta	856.88	26.08
Tayrona	69.06	2.10
Natural Reserve of the Civil Society	**0.08**	**0.00**
las Aves El Dorado	0.01	0.00
Rancho Luna	0.07	0.00
Huitoto	**15,338.72**	**17.91**
Natural National Park	**15,338.72**	**17.91**
Cahuinari	5,591.41	6.53
Rio Pure	9,723.91	11.36
yaigoje Apaporis	23.40	0.03
La Gloria	**4.74**	**0.17**
National Protected Forest Reserve	**4.74**	**0.17**
Cano Alonso	4.74	0.17
Lebrija	**223.79**	**3.45**
Integrated Management Regional District	**223.79**	**3.45**
Humedal San Silvestre	21.23	0.33
Serrania de Los yariguies	202.56	3.12
Macarena	**3,167.15**	**57.88**
Natural National Park	**3,167.15**	**57.88**
Sierra de la Macarena	2,849.20	52.07
Tinigua	317.94	5.81
Macuira	**238.90**	**96.14**
Natural National Park	**238.90**	**96.14**
Macuira	238.90	96.14
Maipures	**5,981.26**	**58.74**
Natural National Park	**2,996.60**	**29.43**
El Tuparro	2,996.60	29.43
UNESCO-MAB Biosphere Reserve	**2,984.66**	**29.31**
El Tuparro	2,984.66	29.31
Marocaso	**136.97**	**10.89**
Integrated Management Regional District	**136.84**	**10.88**
Banaderos	136.84	10.88
Natural Reserve of the Civil Society	**0.13**	**0.01**

San Martin	0.13	0.01
Micay	**2,770.47**	**18.45**
National Protected Forest Reserve	**1,039.64**	**6.92**
Rio Anchicaya	983.77	6.55
Rio Escalarete y San Cipriano	55.87	0.37
Regional Protected Forest Reserve	**6.45**	**0.04**
Serrania Pinche	6.45	0.04
Natural National Park	**1,724.38**	**11.48**
Los Farallones de Cali	1,407.40	9.37
Munchique	316.98	2.11
Murri	**24.86**	**0.25**
National Protected Forest Reserve	**24.86**	**0.25**
Zona Musinga Carauta	24.86	0.25
Nechi	**593.42**	**1.15**
National Protected Forest Reserve	**28.92**	**0.06**
Quebrada Penon	6.42	0.01
Rio Nare	22.50	0.04
Regional Protected Forest Reserve	**93.74**	**0.18**
El Contento las Palmas	0.23	0.00
la Copa San Jose	1.15	0.00
Punchina	37.20	0.07
Reserva Forestal Protectora Regional Cerro Bravo	9.00	0.02
San Lorenzo	46.16	0.09
Integrated Management Regional District	**389.48**	**0.76**
Complejo Cienagoso de Zarate Malibu y Veladero	5.66	0.01
Cuchilla de San Antonio	0.36	0.00
Canon del Rio Alicante	63.53	0.12
Sistema de Paramos y Bosques Altoandinos del Noroccidente Medio Antioqueno	58.51	0.11
Divisoria Valle de Aburra Rio Cauca	90.46	0.18
Embalse El Penol y Cuenca Alta del Rio Guatape	168.22	0.33
laguna del Coco	2.75	0.01
Natural National Park	**76.14**	**0.15**
Selva de Florencia	76.14	0.15
Natural Regional Park	**1.05**	**0.00**
Metropilitano Cerro El Volador	**1.05**	**0.00**
Recreational Area	**0.30**	**0.00**
Parque Ecologico Cerro Nutibara	0.30	0.00
Natural Reserve of the Civil Society	**3.80**	**0.01**

El Garcero	3.80	0.01
Piedemonte Meta	**4.44**	**0.05**
Recreational Area	**1.49**	**0.02**
Parque Ecologico Humedal Calatrava	0.07	0.00
Parque Ecologico Humedal Caracoli	0.33	0.00
Parque Ecologico Humedal Charco Oasis	0.11	0.00
Parque Ecologico Humedal Coroncoro	0.29	0.00
Parque Ecologico Humedal Zuria	0.68	0.01
Soil Conservation District	**2.96**	**0.03**
Suelos Kirpas Pinilla la Cuerera	2.96	0.03
Rio Sucio	**1,650.80**	**16.57**
National Protected Forest Reserve	**352.48**	**3.54**
Darien	6.72	0.07
Rio Leon	345.75	3.47
Natural National Park	**649.77**	**6.52**
Los Katios	648.37	6.51
Paramillo	1.40	0.01
World Heritage Site	**648.56**	**6.51**
Los Katios	648.56	6.51
Sabanas Altas	**5,146.58**	**5.74**
Natural National Park	**2,580.74**	**2.88**
El Tuparro	2,580.74	2.88
Natural Regional Park	**13.10**	**0.01**
Laguna de Lomalinda	8.14	0.01
Laguna San Vicente	4.96	0.01
Natural Reserve of the Civil Society	**10.31**	**0.01**
Ana Maria	0.23	0.00
El Tigrillo	6.32	0.01
la Macarena	3.77	0.00
UNESCO-MAB Biosphere Reserve	**2,542.43**	**2.83**
El Tuparro	2,542.43	2.83
Selvas del Norte de Guaviare	**0.76**	**0.00**
Forest Reserve	**0.76**	**0.00**
Sipapo	0.76	0.00
Sinu-San Jorge	**5,939.83**	**21.56**
Integrated Management Regional District	**1,474.42**	**5.35**
Complejo de Humedales de Ayapel	1,474.42	5.35
Natural National Park	**4,463.56**	**16.20**

Paramillo	4,463.56	16.20
Natural Reserve of the Civil Society	**1.85**	**0.01**
Santa Fe	1.85	0.01
Tacarcuna	**391.65**	**60.83**
National Protected Forest Reserve	**391.65**	**60.83**
Darien	391.65	60.83
Ticuna	**2,606.60**	**28.63**
Natural National Park	**2,606.60**	**28.63**
Amacayacu	2,606.60	28.63
Tumaco	**756.26**	**6.57**
Ecological Reserve	**4.82**	**0.04**
Manglares Cayapas Mataje	4.82	0.04
Natural National Park	**751.45**	**6.53**
Sanquianga	751.45	6.53
Turbo	**344.14**	**4.59**
Integrated Management Regional District	**286.21**	**3.81**
Ensenada de Rionegro	286.21	3.81
Natural National Park	**36.58**	**0.49**
Paramillo	36.58	0.49
Natural Reserve of the Civil Society	**21.35**	**0.28**
Campo Alegre	4.09	0.05
Reserva Natural Horizontes	17.26	0.23
Utria	**72.20**	**1.97**
Natural National Park	**72.20**	**1.97**
Utria	72.20	1.97
Yari-Mariti	**30,399.93**	**44.17**
Indigenous Area	**0.86**	**0.00**
Alto Rio Negro	0.28	0.00
Rio Apaporis	0.59	0.00
Natural National Park	**30,399.07**	**44.17**
Cahuinari	15.61	0.02
Cordillera de Los Picachos	794.70	1.15
Serrania de Chiribiquete	21,424.58	31.13
Sierra de la Macarena	15.43	0.02
Tinigua	1,803.85	2.62
Yaigoje Apaporis	6,344.90	9.22
Grand Total	**164,616.88**	**792.53**

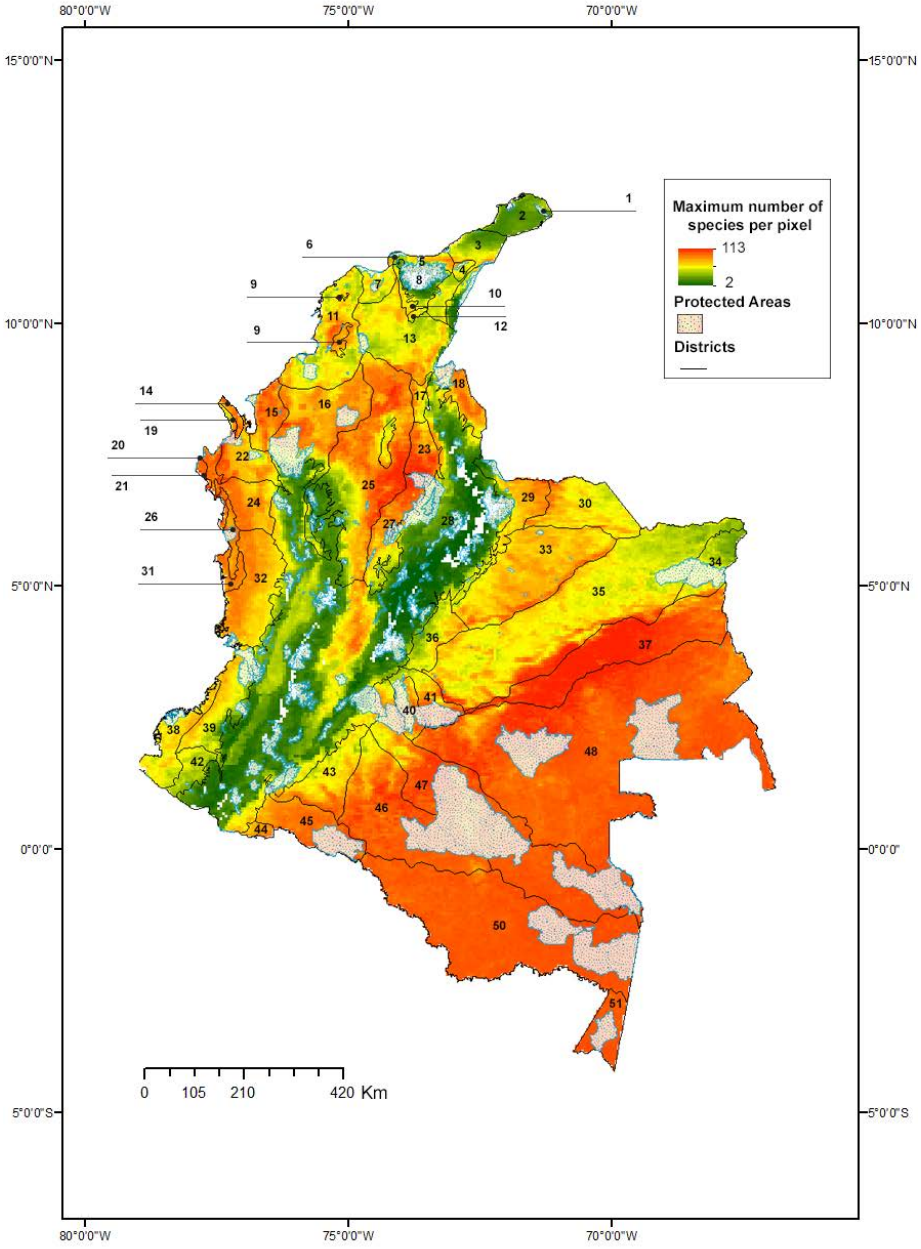


Figure 3.3. Protected areas in Colombia in relation to areas of high predicted species richness in Sapotaceae.
Species richness of neotropical Sapotaceae is represented by colours from red to green. Red colours represent areas of high predicted species richness and green colours areas of low predicted species richness based on SDM results. Protected Areas correspond to those listed by IUCN (2016). Black lines and numbers represent Corzo's (2008) Districts. 1: Macuira, 2: Alta Guajira, 3: Baja Guajira y Alto Cesar, 4: Marocaso, 5: Guachaca, 6: Santa Marta Enclaves Azonales, 7: Delta del Magdalena , 8: Chundua, 9: Maria y Piojo, 10: Aracataca, 11: Cartagena, 12: Caracolicito, 13: Ariguani-Cesar, 14: Acandi-San Blas, 15: Turbo, 16: Sinu-San Jorge, 17: La Gloria, 18: Catatumbo, 19: Tacarcuna, 20: Aspave-El Limon-Pirre, 21: Jurado, 22: Rio Sucio, 23: Lebrija, 24: Murri, 25: Nechi, 26: Baudo, 27: Carare, 28: Perija, 29: Arauca-Apure, 30: Piedemonte Casanare-Arauca, 31: Utria, 32: Alto Atrato-San Juan, 33: Casanare, 34: Maipures, 35: Sabanas Altas, 36: Piedemonte Meta, 37: Selvas del Norte de Guaviare, 38: Tumaco, 39: Micay, 40: Macarena, 41: Ariari-Guayabero, 42: Barbacoas, 43: Florencia, 44: Kofan, 45: Alto Putumayo, 46: Caguan, 47: Yari-Mariti, 48: Complejo Vaupes, 50: Huitoto, 51: Ticuna.

3.4 Discussion

3.4.1 Areas of highest predicted species richness and their current status

Based on the map of predicted species richness of Sapotaceae, the Selvas del Norte de Guaviare, the Complejo Vaupes, and the Nechi Districts are the most species-rich biogeographic units in Colombia of the biogeographic Districts proposed by Corzo (2008). Two of these Districts are found in the Amazon (Selvas del Norte de Guaviare and Complejo Vaupes) and one in the Magdalena valley (Nechi). Both, the Selvas del Norte del Guaviare and the Complejo Vaupes Districts are located in the Amazon close to Los Llanos, and are in fact neighbouring Districts, highlighting that the general area of these two Districts is species rich. High values in species richness within Selvas del Norte del Guaviare and some parts of the Complejo Vaupes may be the result of the confluence between elements from the lowland rain forests of the Amazon with the savannahs of Los Llanos. Edaphic variation in Selvas del Norte del Guaviare and Complejo Vaupes may also be influencing high values of species richness (Antonelli and Sanmartín, 2011).

Selvas del Norte del Guaviare is considered to be under-represented in the Colombian system of protected areas with only one Forest Reserve covering <1% of its total extension. In Complejo Vaupes, six protected areas are found, but none of them cover the northern part of the District where high species richness is predicted. This could be a consequence of the previous lack of data on species distributions of plants, and of previous assumptions in which large protected areas in other parts of the Amazon were thought to represent the region as a whole.

High predicted species richness was also found in the Nechi District in the Magdalena Valley. This District is located between the Eastern and Western Cordilleras in a region known to host high endemism, and many timber species with high economic value (e.g. *Isidodendron tripterocarpum*; Fernández-Alonso *et al.,* 2007). The biotic richness of this region may have originated before the rise of the Eastern Cordillera in Colombia, and when the Magdalena valley forests were part of a migration route between taxa from forests to the east (e.g., Amazon and Atlantic Forest) and west (e.g., Choco) of the Eastern Cordillera (Chapter 2 and Bernal *et al.,*

2016). The highest species richness in Sapotaceae within Nechi is found towards the eastern part of the District. Eighteen Protected Areas have been established in Nechi but as in the Complejo Vaupes District, none of them are located in areas of potential high richness of Sapotaceae species. This is because much of the original forest has been cleared or severely modified/disturbed, and the remaining small forest areas cannot be assigned as protected areas, because they do not meet criteria related to the total area needed to host communities of plant and animal taxa with long-term stability. The absence of protected areas in the lowland rain forest of the Magdalena valley is also the result of the lack of data on species distributions and diversity due to the restricted access to these forests that in the past were a base for illegal armed groups.

3.4.2 Future directions for conservation in Colombia

The civil conflict in Colombia has been an obstacle not only for the social and political development of the country, but it has also prevented the development of biological studies. Illegal armed groups were traditionally hidden in some of the most inaccessible regions of Colombia, areas that could potentially host the most species-rich and distinctive biotas, such as the forests of the Magdalena valley.

In November of 2016 this changed as one of the biggest illegal armed groups, the FARC, signed a peace agreement with the Colombian government. This Act has recently been followed by formal communications with other revolutionary groups, who have announced their intention to also be part of the peace process and to end the civil conflict. The long awaited arrangements between rebels and the Colombian government mean that illegal armed group camps will no longer be concentrated within Colombian forests, opening paths for economic growth, for the social development of communities previously under the enforced laws of illegal groups, and opening opportunities for scientific research (Clerici *et al.,* 2016; Wade, 2017).

The study of these newly accessible ecosystems, including areas of lowland rain forests, must now be a priority for scientists and policy makers, to ensure that new open access to areas rich in biotic resources and the arrival of foreign and local investors is in accord with conservation priorities, and does not result in an increase in biodiversity loss. These studies should incorporate an update on the delimitation of biogeographic Districts, including those proposed by Corzo (2008) which were found

to rely on physical barriers such as mountain ranges and rivers, and not to depict patterns of predicted richness in tree species such as Sapotaceae. An update on these units will likely facilitate the achievement of conservation goals in Colombia, and it could guarantee the preservation of ecosystems of high biotic interest.

The preservation of ecosystems of high biotic interest in Colombia could also be aided by future economic investments and an increase in the representation of the northern Amazon and the Magdalena inter-Andean valley in the National system of Protected Areas. Although endemic species in Sapotaceae were not found within those units in Colombian forests, the northern Amazon and the Magdalena inter-Andean valleys could host biotas of high conservation value based on the occurrence of unique assemblages of plant species. Other taxonomic groups could reveal additional patterns of endemism in Colombia (Gentry, 1986; Bernal *et al*., 2016).

Analyses of distribution patterns using predicted species richness based on SDMs as a proxy to understand the dynamics of the lowland rain forest biome are a step forward to achieve conservation goals. Measures of species richness at national scales like those presented here could complement previous work where the main criteria has been the occurrence of endemics and habitat degradation (e.g., biodiversity hotspots). The use of SDMs takes advantage of the available data on species occurrence, crucial in a scenario where 100% inventories, and complete floras of all ecosystems will not be accessible in the short term, and where the current rates of biodiversity loss demand prompt action. Furthermore, the use of species richness as a measure of conservation value facilitates the identification of Protected Areas based on the "minimum set criterion" or where the maximum number of species is preserved within the smallest possible extension (e.g., Pawar *et al.,* 2007). The establishment of large Protected Areas is not always feasible, and is limited by national budgets where economic resources for conservation are not priority.

3.4.3 Future research priorities

It should be noted, however, that species richness is only one of the possible criteria that could be used to identify areas of high conservation value, and although analyses like those presented in this work add to a general understanding on conservation in the Neotropics, they should be complemented by further work where

other measures are considered. Further steps could include analyses of distinctiveness of each District/region in terms of species composition, analyses of distribution patterns of other representative groups, and field verification and data collection in the areas that have shown high predicted species richness.

Complementarity analyses would also be needed to understand the uniqueness of the species protected under each current area for conservation (e.g. national parks) and the complementarity of these in relation to each other. Such analyses would help to further understand the need for potential new protected areas.

Lastly, analyses of species richness as measured by phylogenetic diversity would allow the identification of unique areas in terms of the evolutionary lineages they hold. Such work would require a better understanding of taxonomic species units in relation to phylogenetic lineages, a problem that currently prevents analyses of phylogenetic diversity in Sapotaceae. More baseline taxonomic work, careful herbarium studies using morphological characters combined with densely-sampled molecular phylogenetic analyses would be needed to resolve this. Meanwhile, the patterns of species diversity as highlighted here for the lowland rain forest biome allow us to integrate scientific studies to applied conservation, and to take first steps in evaluating conservation strategies in areas under high threats of habitat degradation, like the Colombian forests.

References

Antonelli, A. & Sanmartín, I. 2011. Why are there so many plant species in the Neotropics? Taxon. 60: 403-414.

Bernal, R., Gradstein, S. R. & Celis, M. 2016. Catálogo de plantas y líquenes de Colombia. Universidad Nacional de Colombia, Sede Bogotá, Facultad de Ciencias, Instituto de Ciencias Naturales. 3060.

BirdLife International. 2017. Important bird areas factsheet: Parque Nacional Natural Sierra de la Macarena.

Boyce, M. S., Vernier, P. R., Nielsen, S. E. & Schmiegelow, F. K. A. 2002. Evaluating resource selection functions. Ecological Modelling. 157: 281-300.

Burnham, R. J. & Johnson, K. R. 2004. South American paleobotany and the origins of neotropical rainforests. Philosophical Transactions of the Royal Society of London, series B. 359: 1595-1610.

Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., … Forzza, R. C. 2017. Amazon plant diversity revealed by a taxonomically verified species list. Proceedings of the National Academy of Sciences. 114: 10695-10700.

CDB (Convenio de Diversidad Biológica). 2004. Programa de trabajo sobre áreas protegidas PNUMA. Montreal. Québec, Canada.

Clerici, N., Richardson, J. E., Escobedo, F. J., Posada, J. M., Linares, M., Sanchez, A., & Vargas, J. F. 2016. Colombia: Dealing in conservation. Science. 354:190.

Chaves, M. & Arango (ed.). 1997. Informe nacional sobre el estado de la Biodiversidad-Colombia. Tomo I. causas de la perdida de la biodiversidad. Instituto investigaciones de recursos biológicos Alexander von Humboldt, PNU-Ma y Monisterio del Medio Ambiente, Bogota.

Corzo, G. 2008. Áreas prioritarias para la conservación “in situ” de la biodiversidad continental en Colombia.

Elith, J., Graham, C. H., Anderson, R. P., Dudık, Lohmann, L. G., Loiselle, B. A… & Zimmermann, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 29: 129-151

Elith, J., Phillips, S. J.., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions. 17: 43-57.

Fondo para la protección del medio ambiente FEN. 1993. Colombia Pacifico. Tomo I. Santa Fe de Bogotá, Colombia.

Fernández-Alonso J. L., Pérez-Zabala J. A. & Idarraga-Piedrahita A. 2007. (ed.). "*Isidodendron tripterocarpum*" Catálogo de la Biodiversidad de Colombia.

Forero, E. 1988. Botanical Exploration and Phytogeography of Colombia: Past, Present and Future. Taxon. 37: 561-566.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., … Michaelsen, J. 2015. The climate hazards infrared precipitation with stations: a new environmental record for monitoring extremes. Scientific Data. 2:150066.

GBIF.org (29th February 2016) GBIF Occurrence Download.

Gentry, A. 1986. Species richness and floristic composition of Choco region plant communities. Caldasia. 15: 71-91.

Hernández, J. I., Hurtado, A., Ortiz, R. & Walschburger, T. 1992. unidades biogeográficas de Colombia. En. G. Halffter (ed.). La Diversidad Biológica en Iberoamérica. Acta Zoológica Mexicana. Volumen Especial. 105-152.

Hirzel, A. H., Lay, G. L., Helfer, V., Randina, C. & Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling. 199: 142-152.

IDEAM, IGAC, IAvH, Invemar, I. Sinchi e IIAP. 2007. Ecosistemas continentales, costeros y marinos de Colombia. Instituto de Hidrología, Meteorología y Estudios Ambientales, Instituto Geográfico Agustín Codazzi, Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Instituto de Investigaciones Ambientales del Pacífico Jhon von Neumann, Instituto de Investigaciones Marinas y Costeras José Benito Vives De Andréis e Instituto Amazónico de Investigaciones Científicas Sinchi. Bogotá, D. C. 276.

IUCN, UNEP-WCMC. 2016. The World Database on Protected Areas (WDPA).

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schroder, B., Lindenborn, J., Reinfelder, V., … Wilting, A. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Diversity and Distributions. 19: 1366-1379.

Li, W. & Guo, Q. 2013. How to assess the prediction accuracy of species presence-absence models without absence data? Ecography. 36: 788-799.

Londoño-Vega, A. C., & Alvarez-Dávila, E. 1997. Composicion floristica de dos bosques (tierra firme y varzea) en la region de Araracuara, Amazonia colombiana. Caldasia. 19: 431-463.

Morrone, J. 2001. Biogeografia de America Latina y el Caribe. First ed. Zaragoza: CYTED, ORCYT-UNESCO, Sociedad Entomológica Aragonesa (SEA).

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403: 853-858.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., … Kassem, K. R. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience. 51: 933-938.

Padme.rbge.org.uk. Royal Botanic Garden Edinburgh. 22 Jun 2016 <www.padme.rbge.org.uk>.

Pawar, S., Koo, M. S., Kelley, C., Ahmed, M. F., Chaudhuri, S., & Sarkar, S. 2007. Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. Biological Conservation. 136: 346-361.

Phillips, S. J., Anderson, R. P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling. 190: 231-259.

Phillips, S. J. & Dudík, M. 2008. Modeling species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography. 31: 161-175.

Smith, A. B. 2013. On evaluating species distribution models with random background sites in place of absences when test presences disproportionately sample suitable habitat. Diversity and Distributions. 19: 867-872.

Rangel-Ch., J. O. 2008. La vegetación de la región amazónica de Colombia - Aproximación inicial-. In: J. O. Rangel-Ch. (ed.). Colombia diversidad biótica VII. Vegetación, palinología y paleoecología de la Amazonia colombiana. Universidad Nacional de Colombia-Instituto de Ciencias Naturales. Bogotá. 1-53.

Reynel, C., Pennington, T. & Särkinen, T. 2013. Cómo se formó la diversidad ecológica del Perú. Lima, Perú. 412.

Särkinen, T., Iganci, J. R. V., Linares-Palomino, R., Simon, M. F., & Prado, D. E. 2011. Forgotten forests - issues and prospects in biome mapping using seasonally dry tropical forests as a case study. BMC Ecology. 11:27.

ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomao, R. P., Guevara, J. E., … Silman, M. R. 2013. Hyperdominance in the Amazonian Tree Flora. Science. 342: 6156.

Tropicos.org. Missouri Botanical Garden. 16 Jun 2016 <http://www.tropicos.org>.

Wan, Z.M. 2014. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sensing of Environment. 140: 36-45.

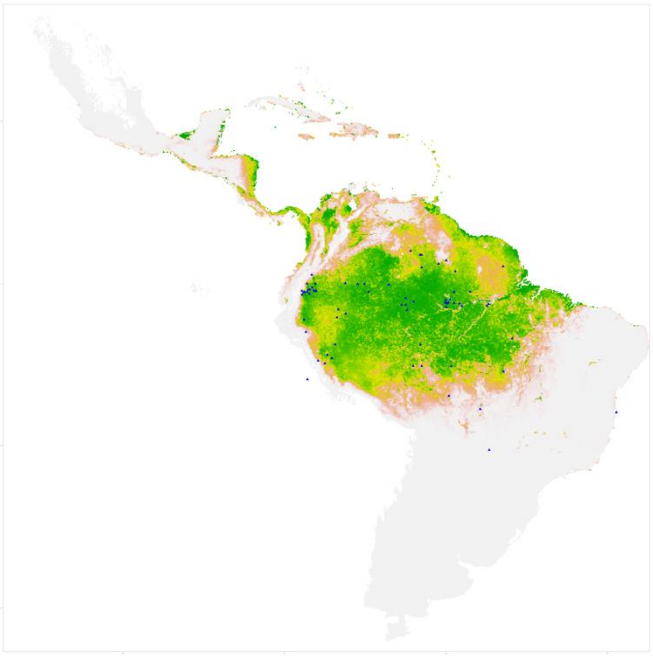
Wan, Z. M. & Dozier, J. 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. IEEE Transactions on Geoscience and Remote Sensing. 34: 892-905.

Wade, L. 2017. In Colombia, peace dividend for science proves elusive. Science. 357:958.

Wiegand, T. & Moloney, K. A. 2014. Handbook of spatial point pattern analysis in ecology. Chapman and Hall/CRC Press, Boca Raton, FL, USA. 538.

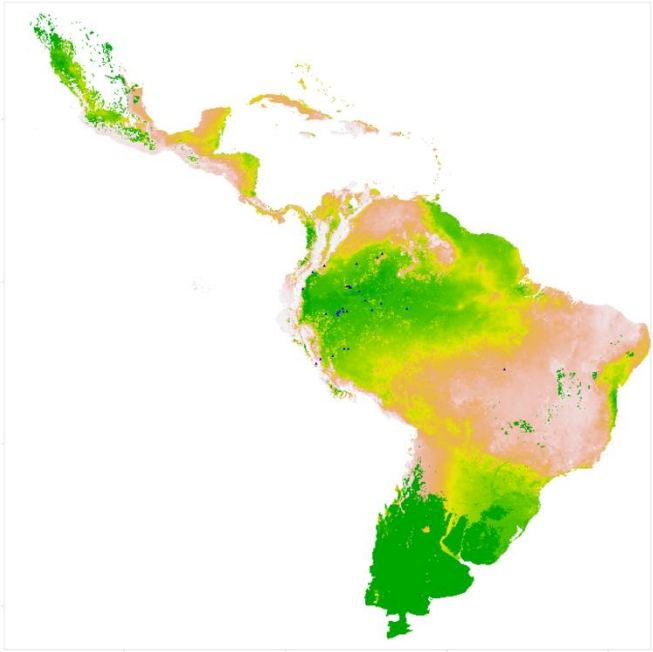
Appendices

Appendix 3.1



Chrysophyllum amazonicum

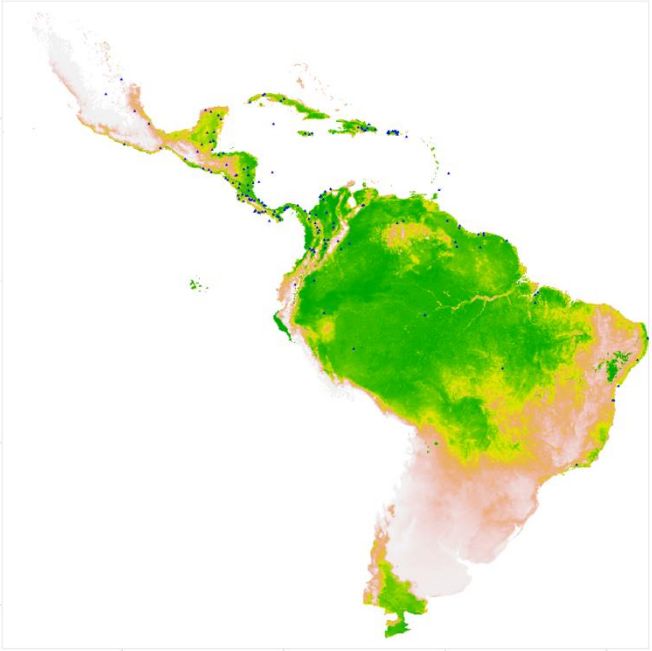
Chrysophyllum argenteum



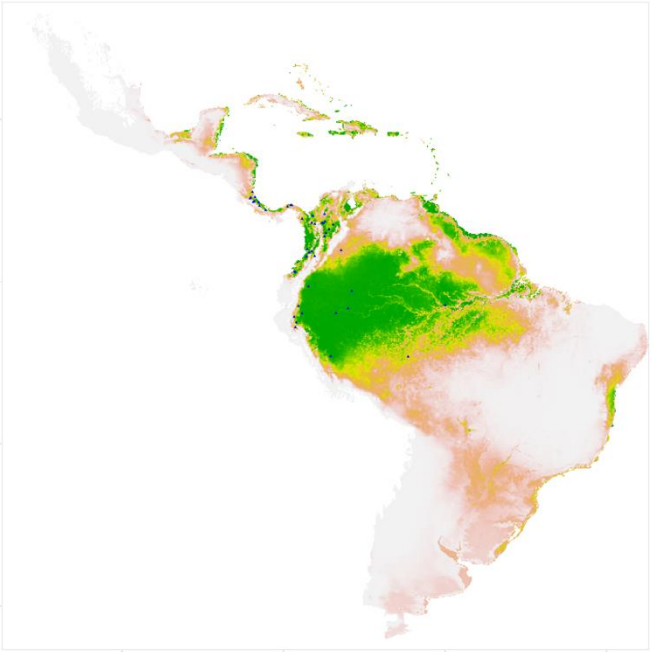
Chrysophyllum *bombycinum*

Chrysophyllum brenesii

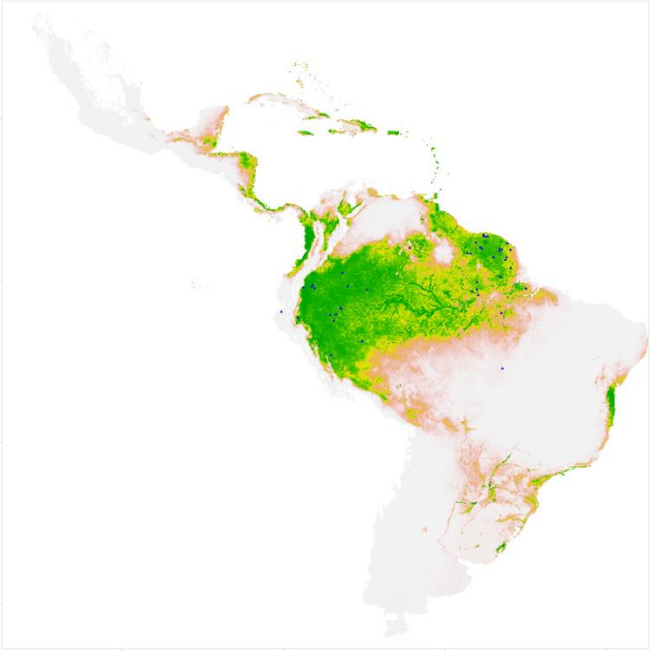
Chrysophyllum caimito



Chrysophyllum cainito

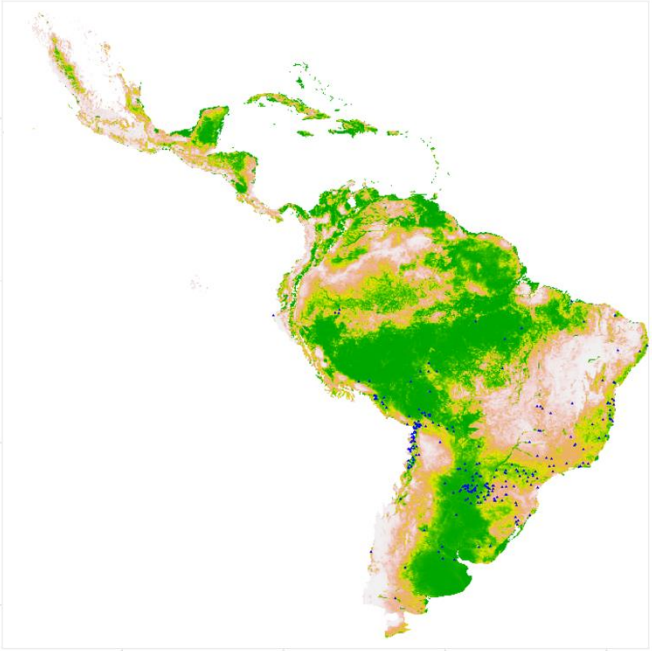


Chrysophyllum colombianum

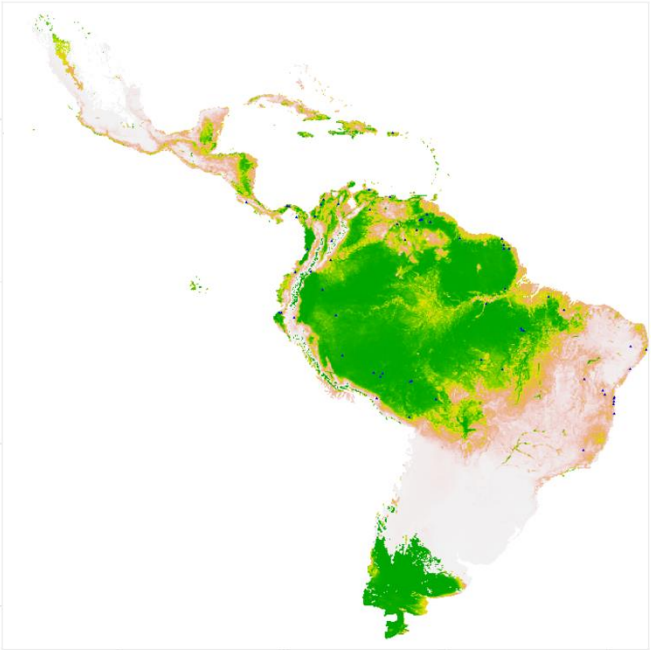


Chrysophyllum cuneifolium

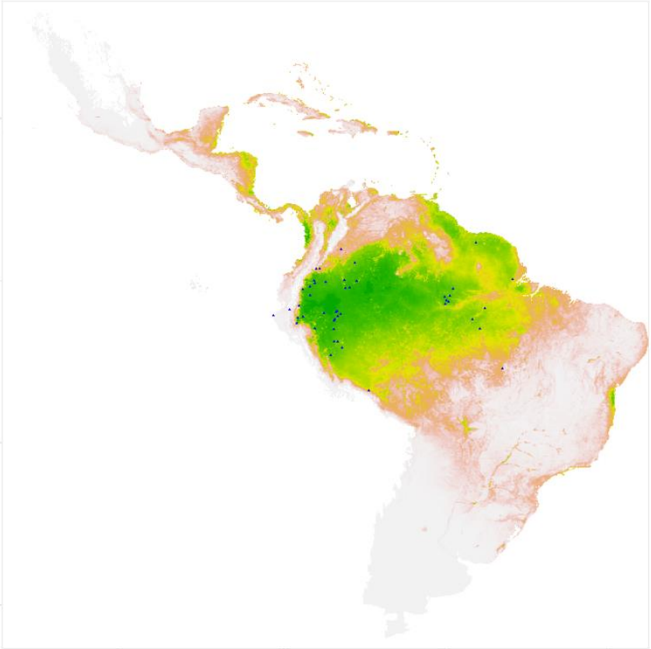
Chrysophyllum flexuosum



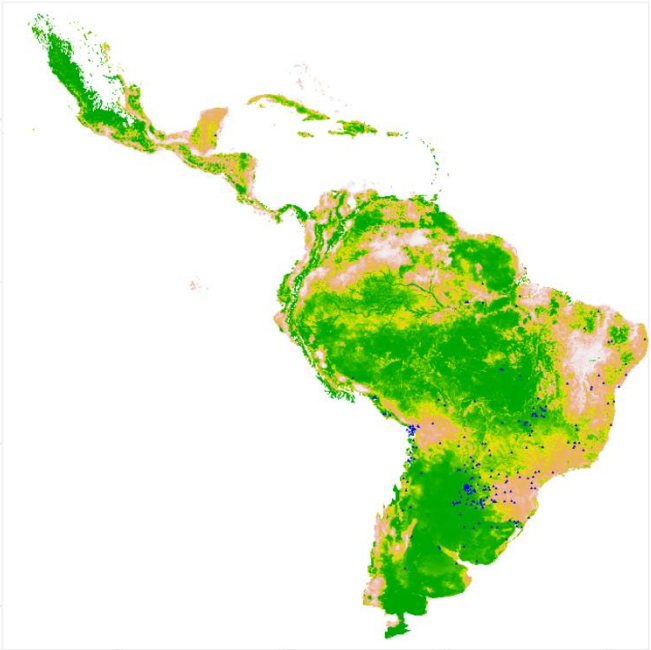
Chrysophyllum gonocarpum



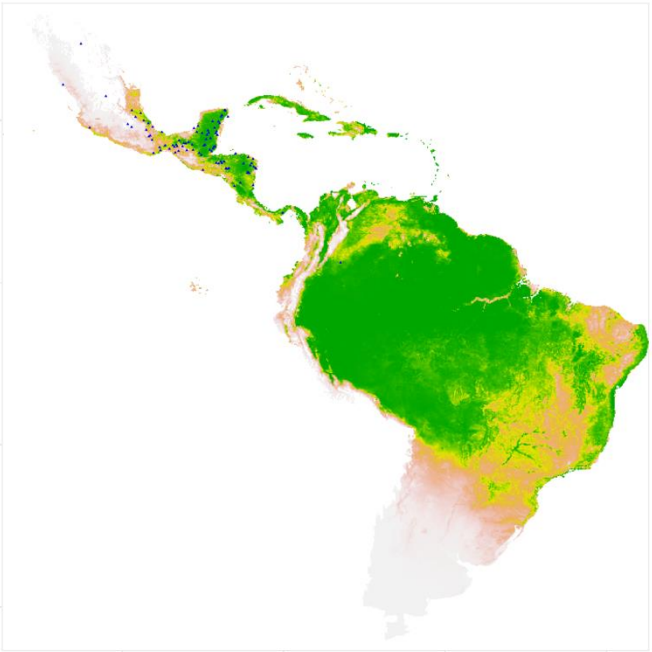
Chrysophyllum lucentifolium



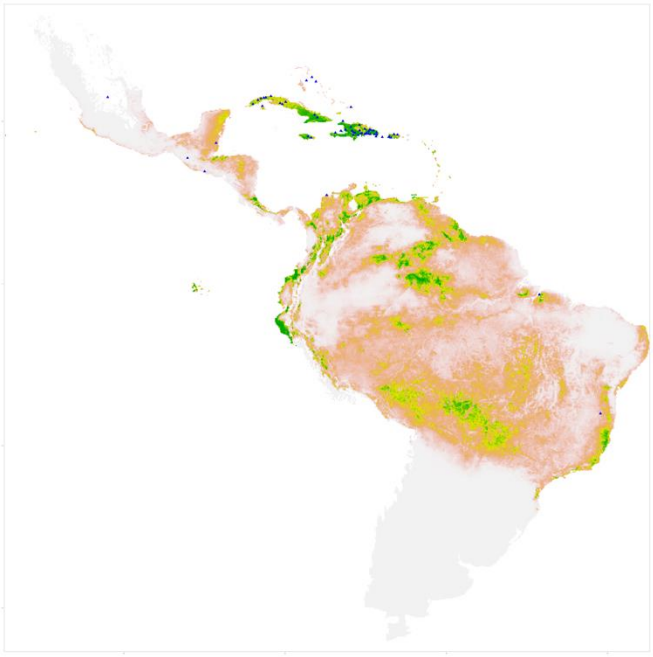
Chrysophyllum manaosense



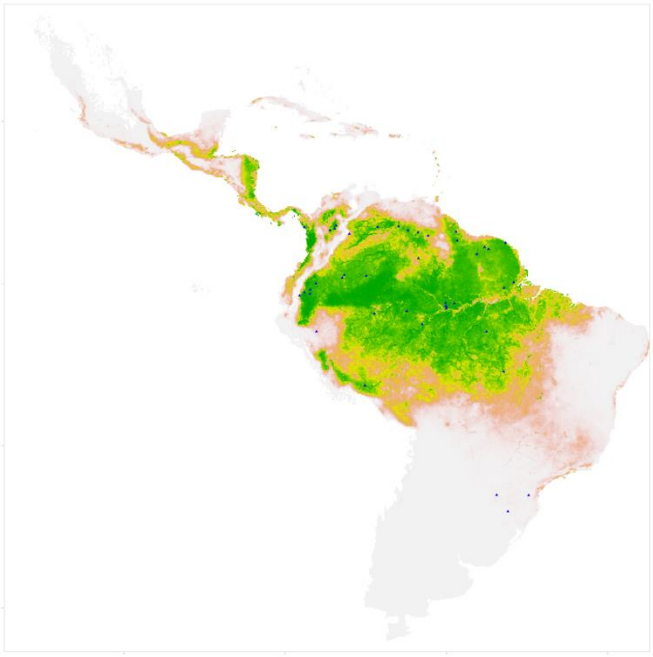
Chrysophyllum marginatum



Chrysophyllum mexicanum`

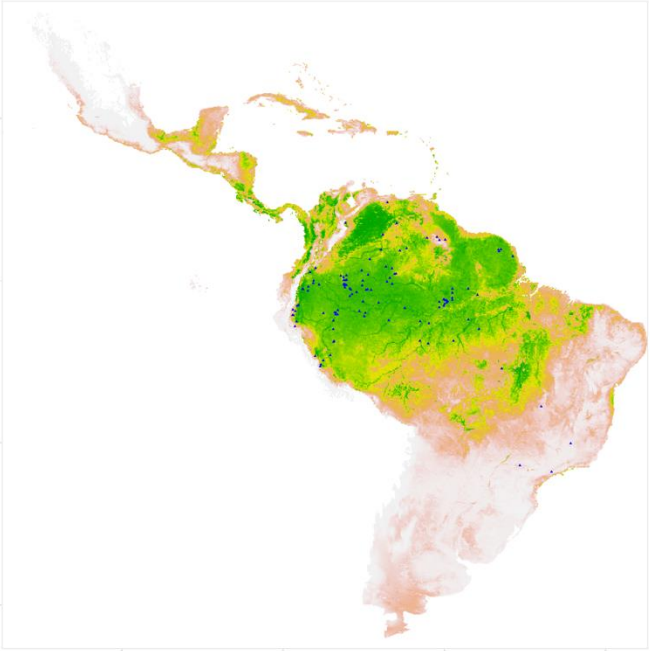


Chrysophyllum oliviforme



Chrysophyllum pomiferum

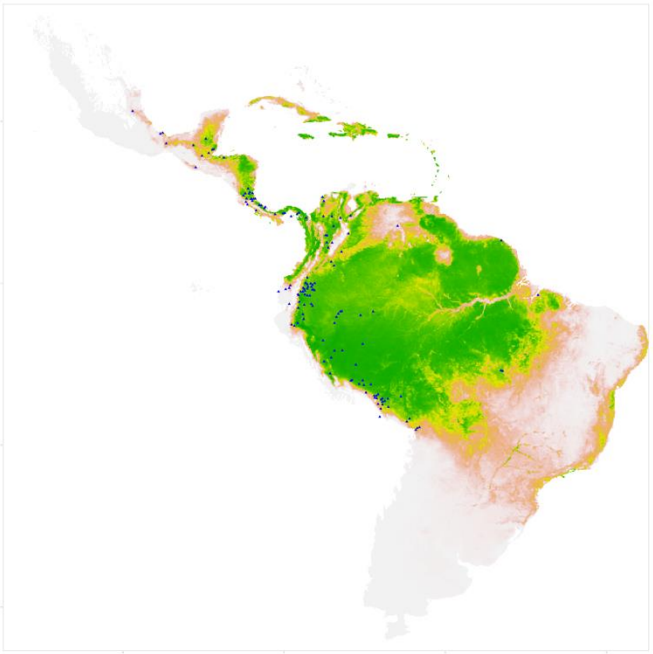
Chrysophyllum prieurii



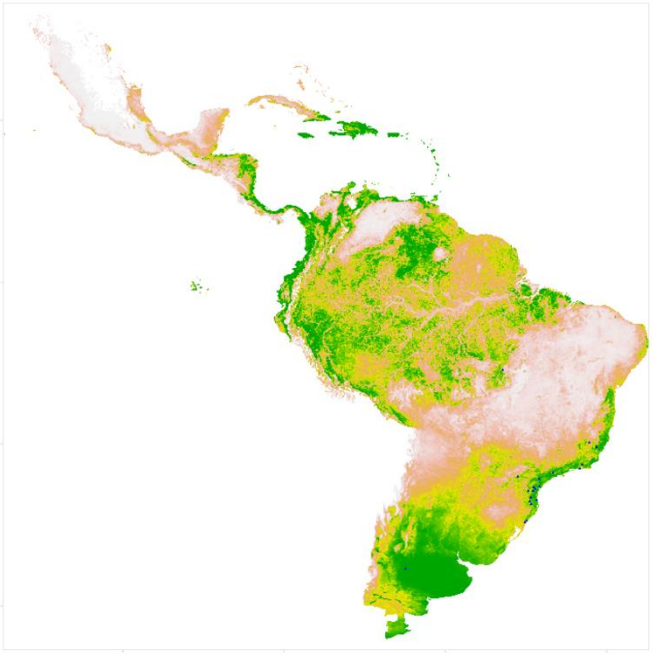
Chrysophyllum sanguinolentum

Chrysophyllum sparsiflorum

Chrysophyllum splendens

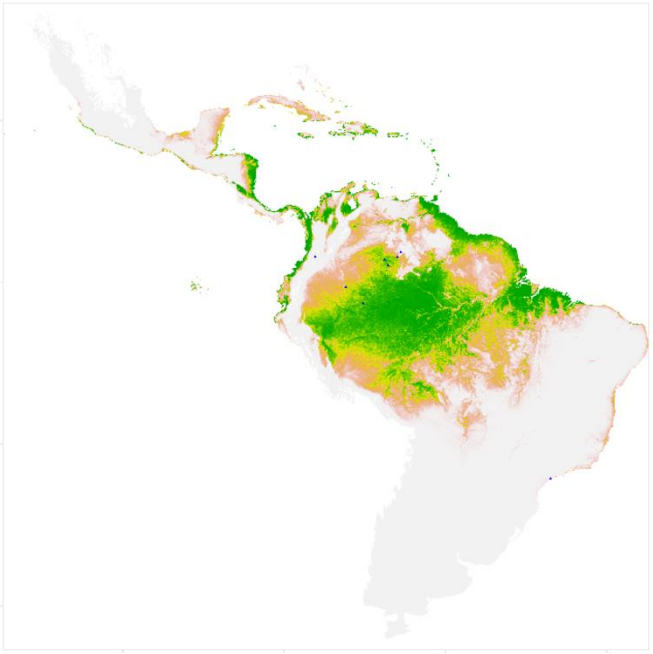


Chrysophyllum venezuelanense

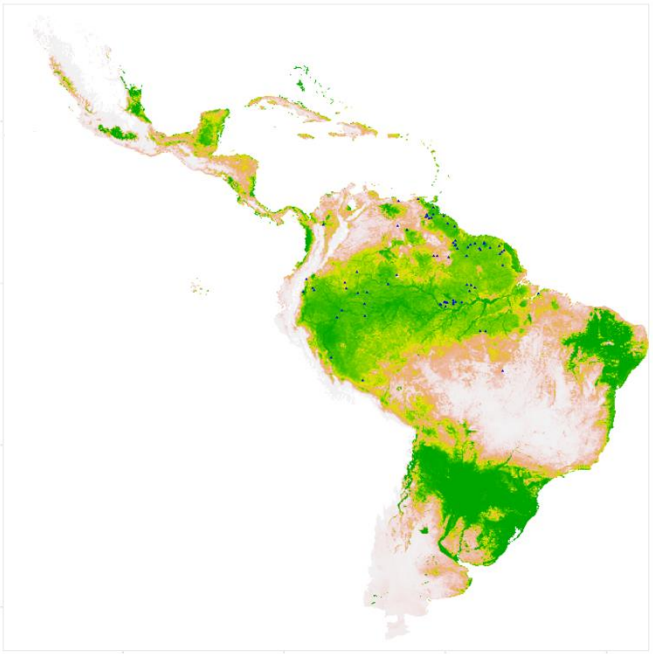


Chrysophyllum viride

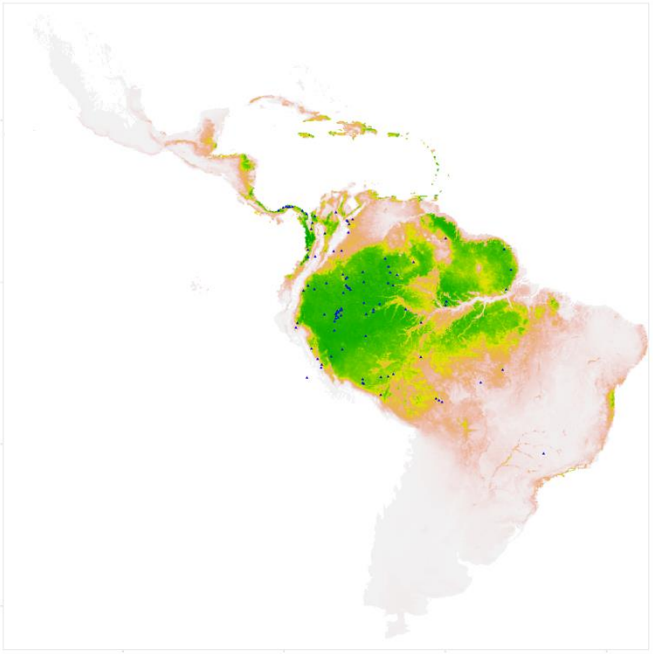
Diploon cuspidatum



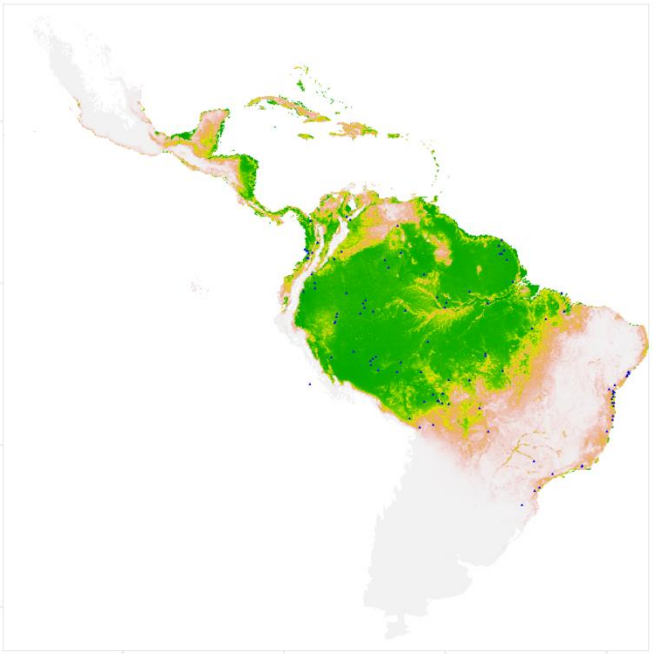
Ecclinusa bullata



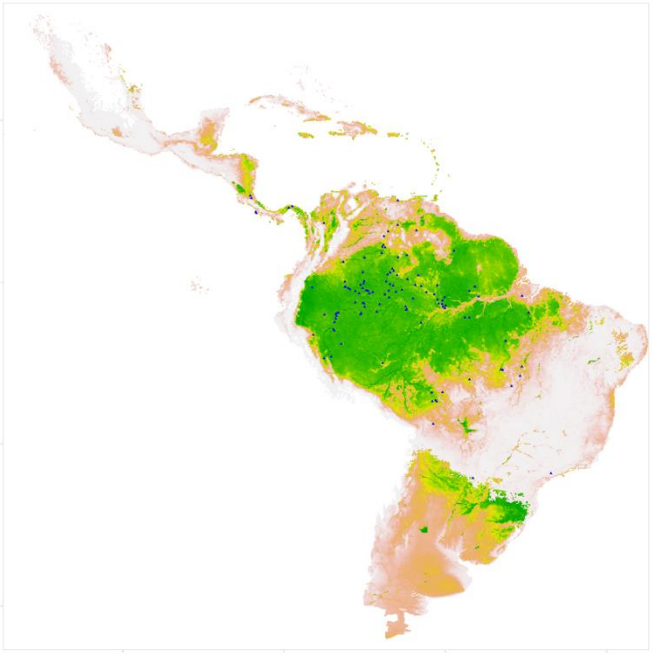
Ecclinusa guianensis



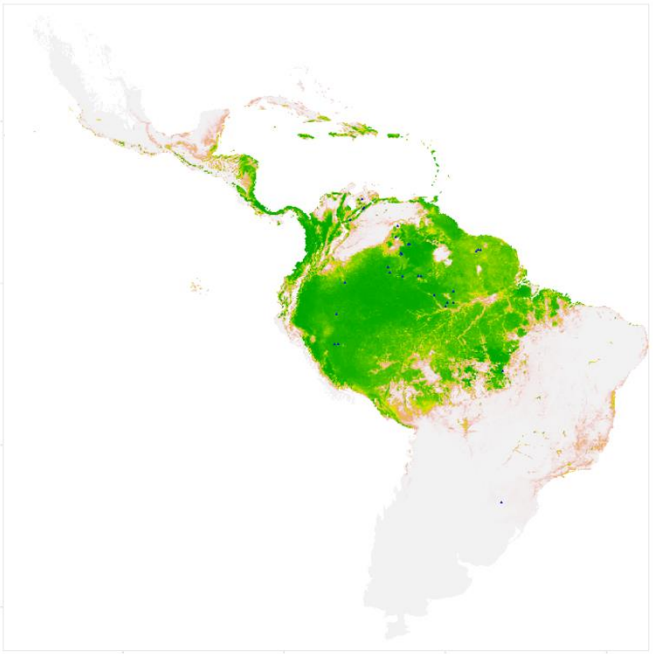
Ecclinusa lanceolata



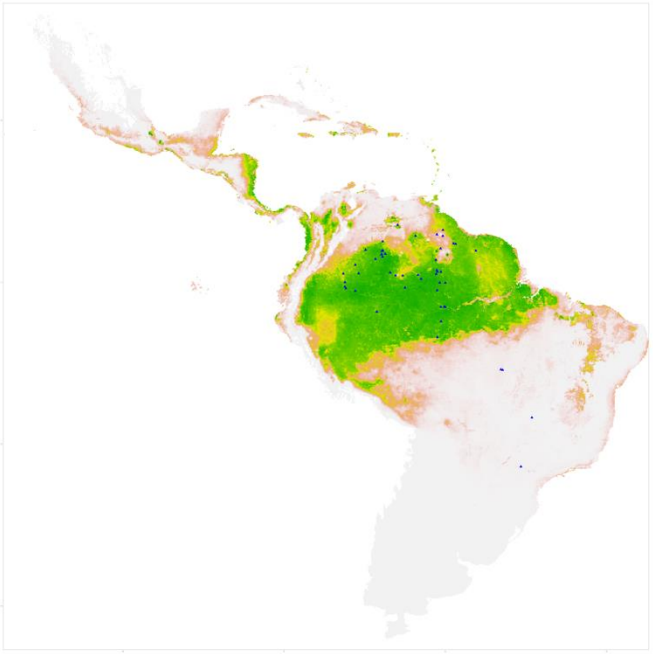
Ecclinusa ramiflora



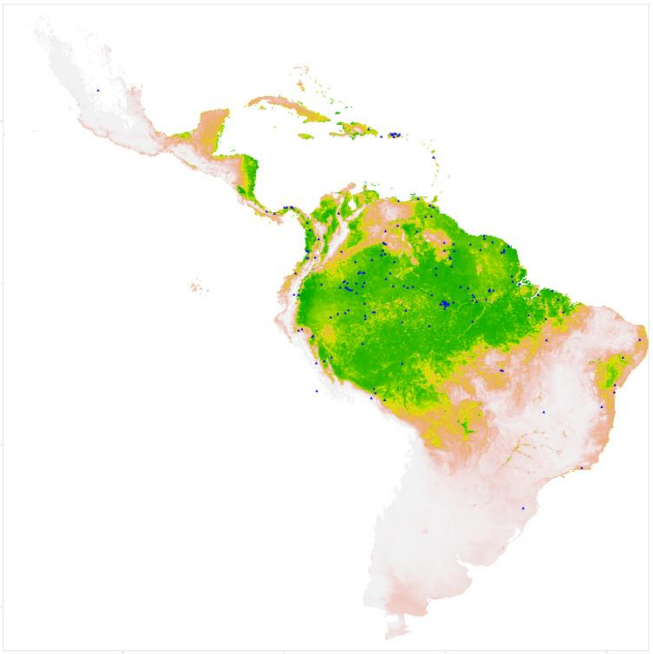
Elaeoluma glabrescens



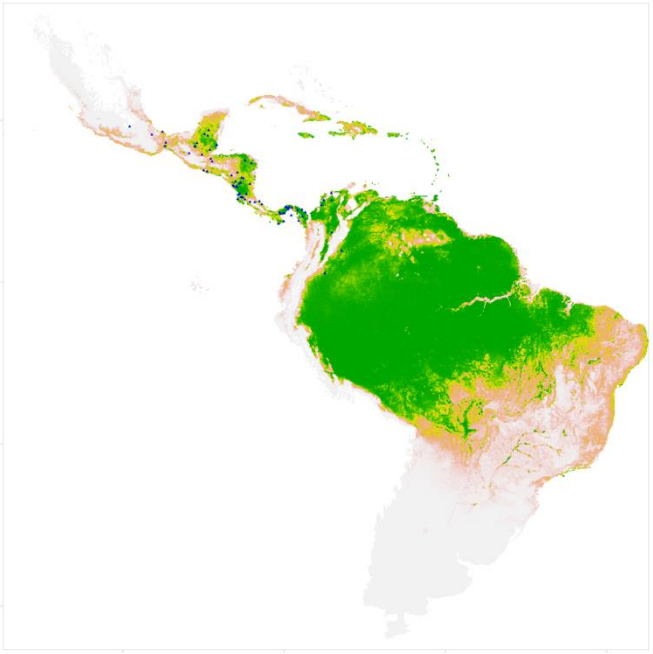
Elaeoluma nuda



Elaeoluma schomburgkiana

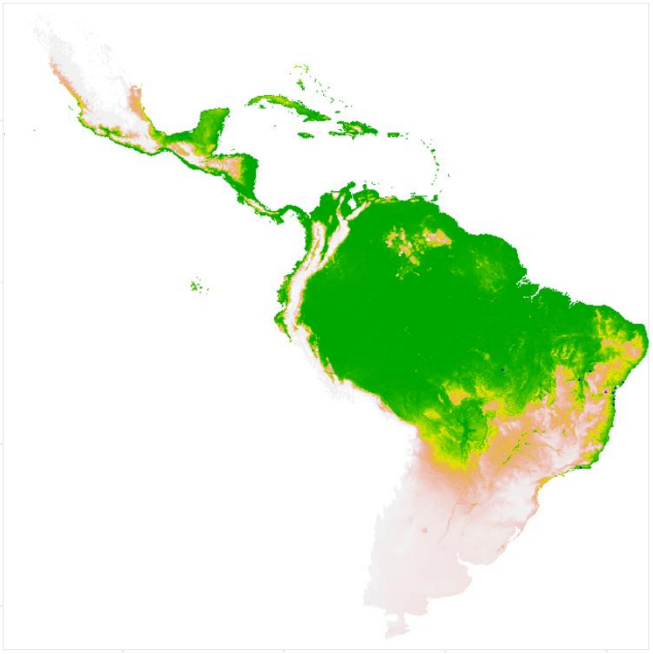


Manilkara bidentata

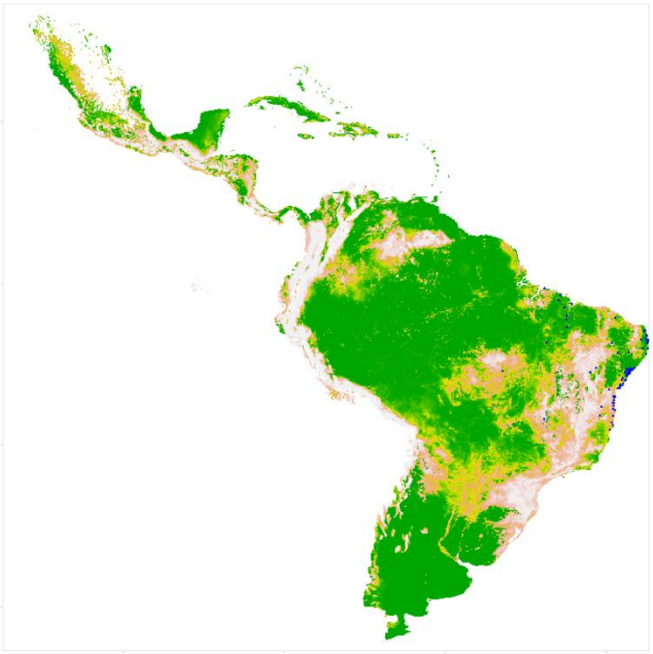


Manilkara chicle

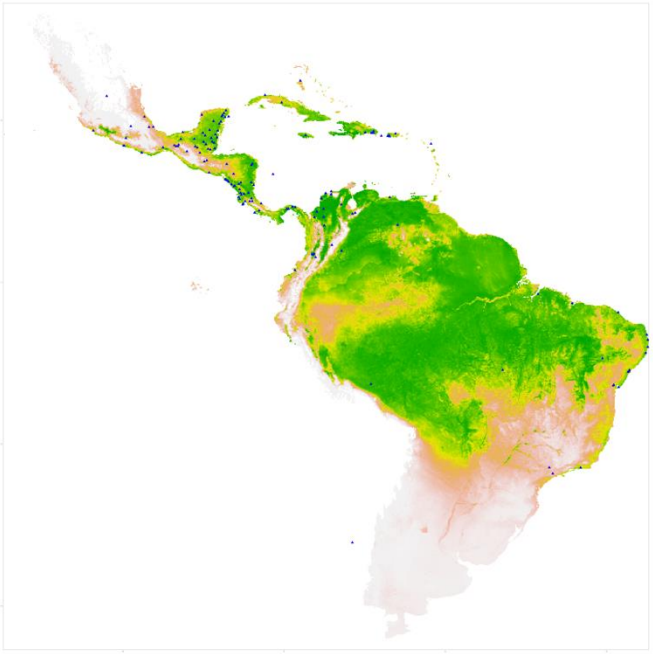
Manilkara huberi



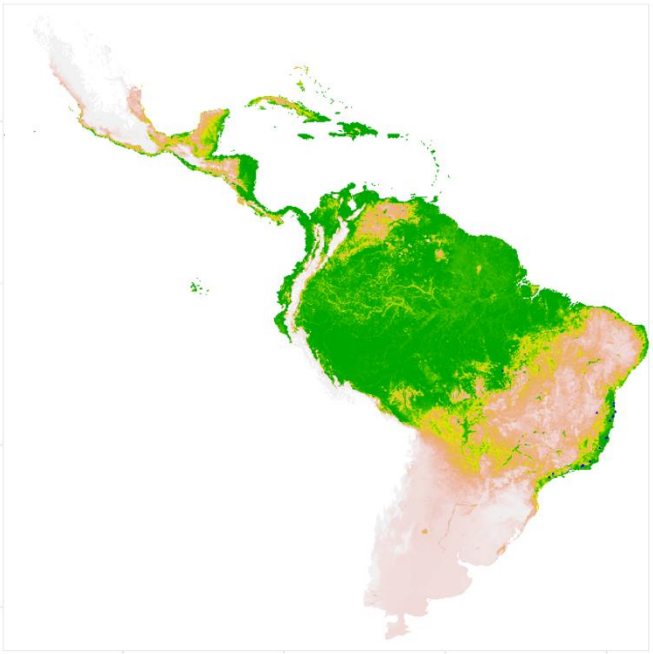
Manilkara longifolia



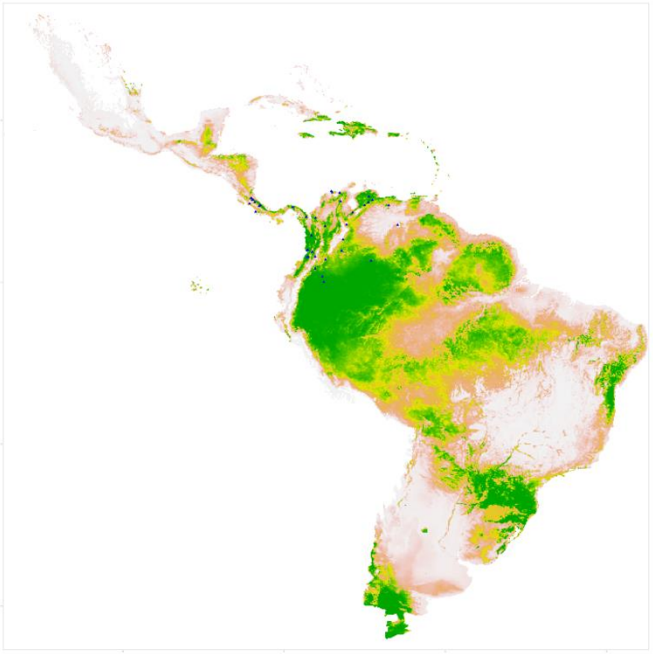
Manilkara salzmannii



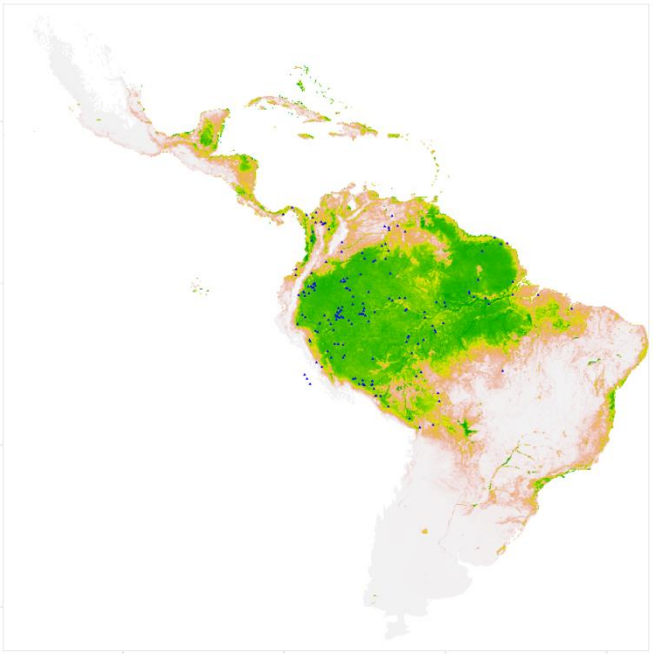
Manilkara zapota



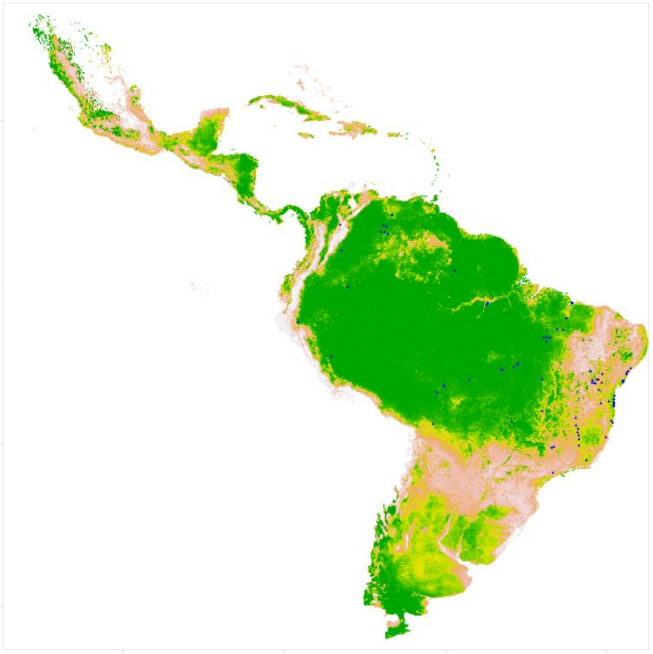
Micropholis crassipedicellata



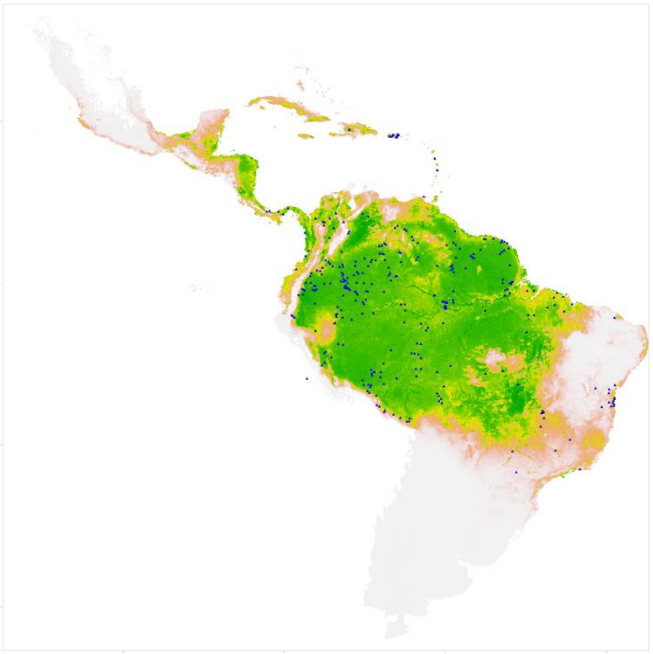
Micropholis crotonoides



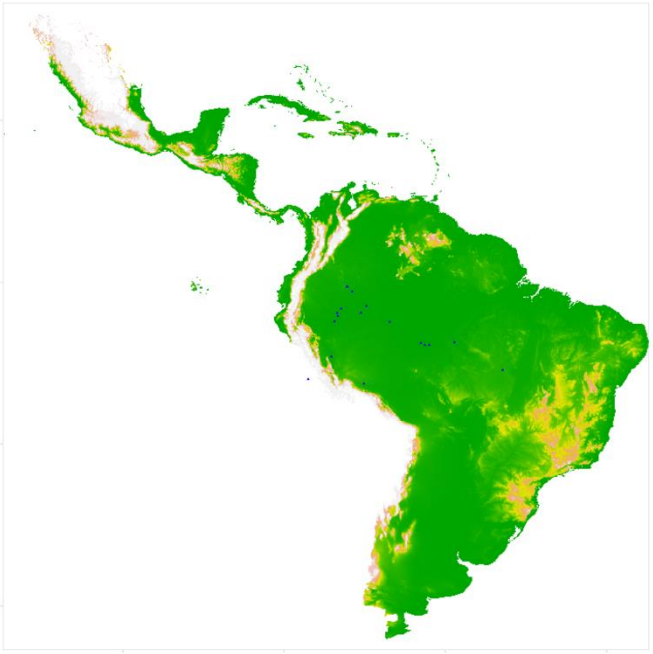
Micropholis egensis



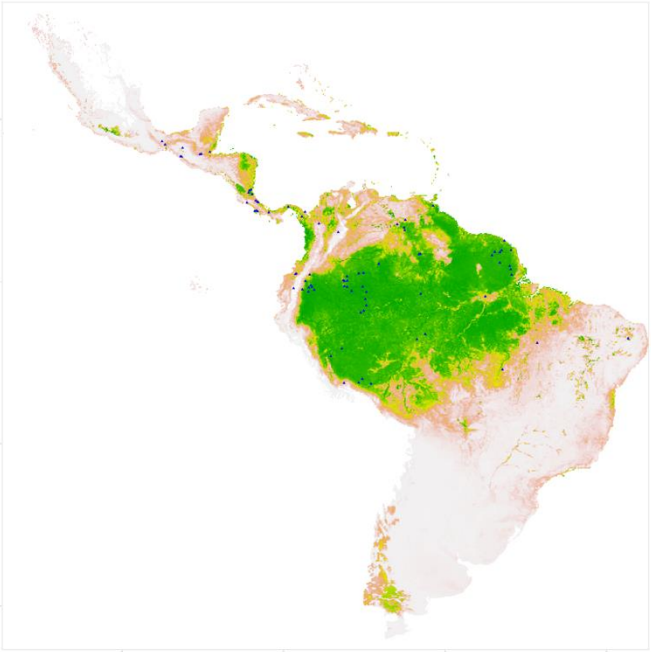
Micropholis gardneriana



Micropholis guyanensis

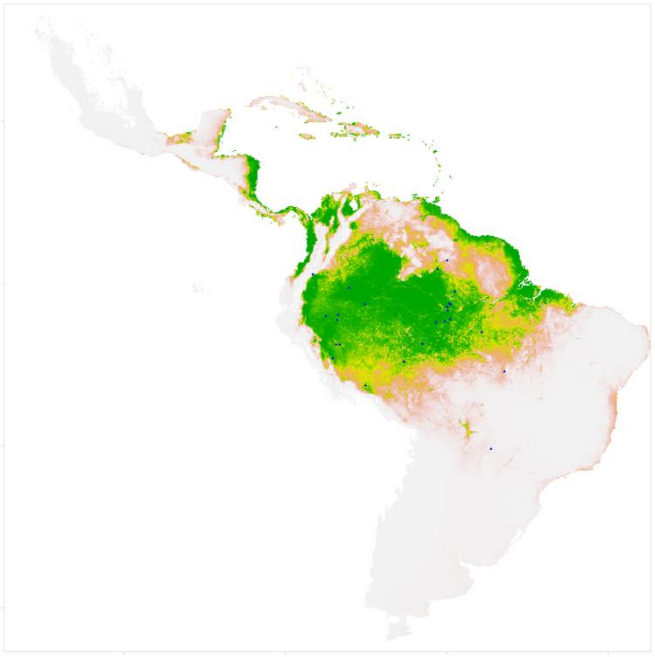


Micropholis madeirensis

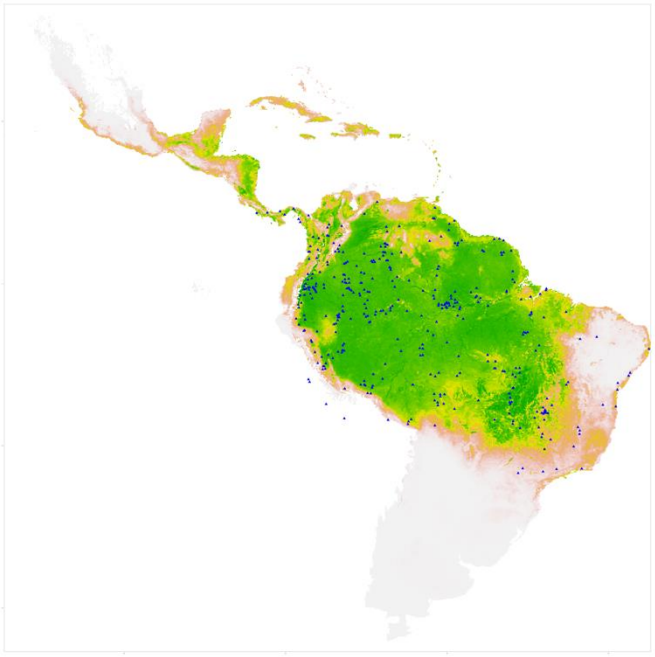


Micropholis melinoniana

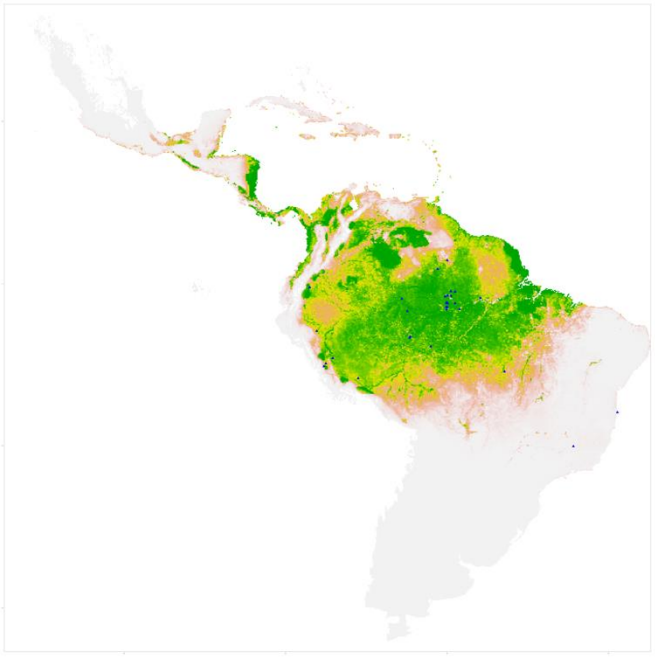
Micropholis porphyrocarpa



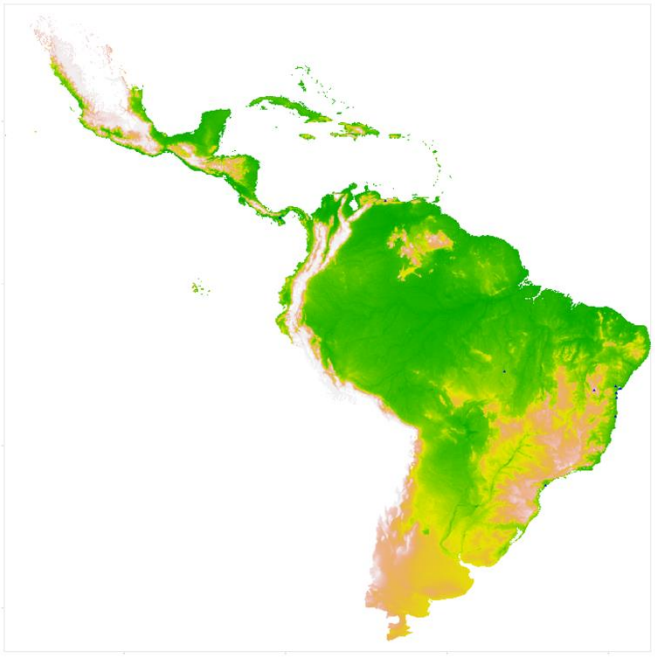
Micropholis trunciflora



Micropholis venulosa

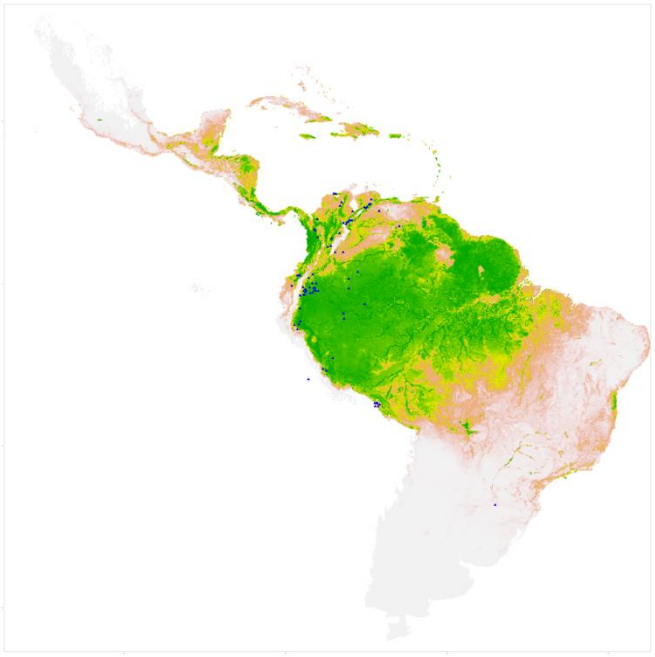


Micropholis williamii

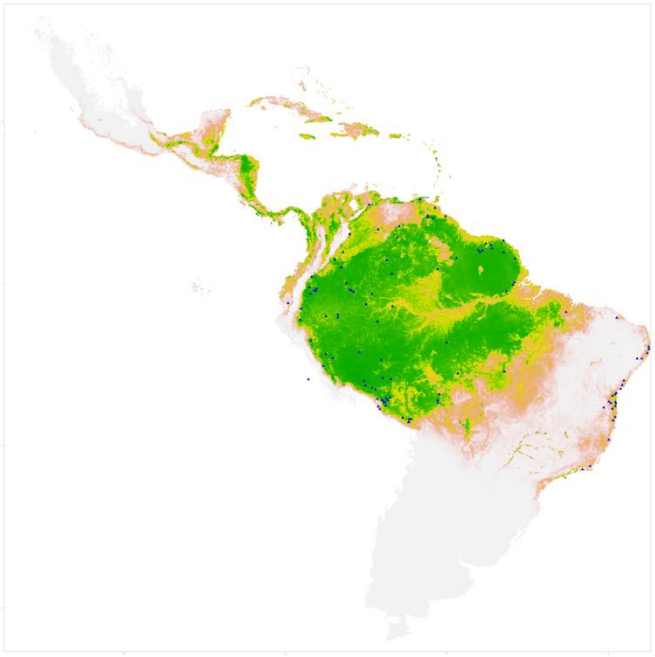


Mimusops coriacea

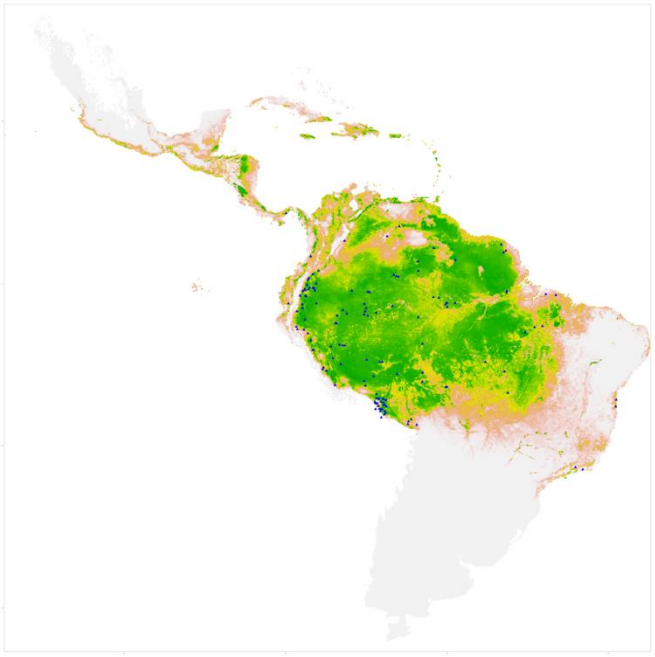
Pouteria ambelaniifolia



Pouteria baehniana

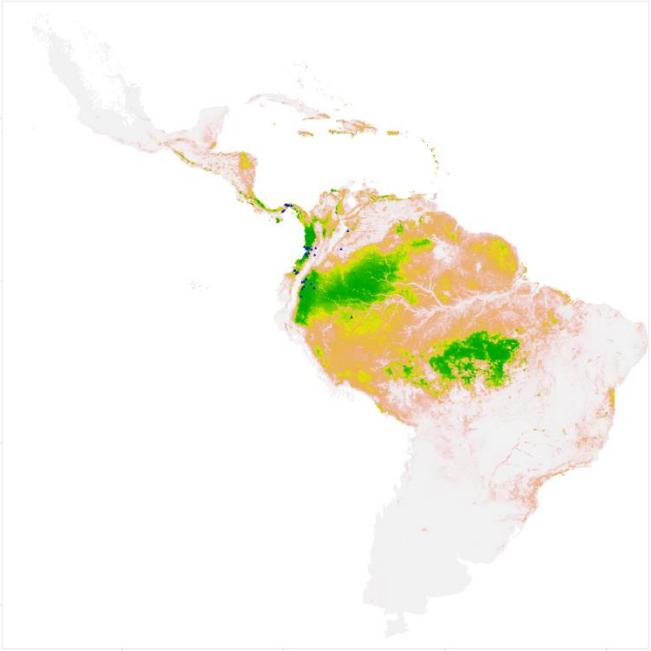


Pouteria bangii

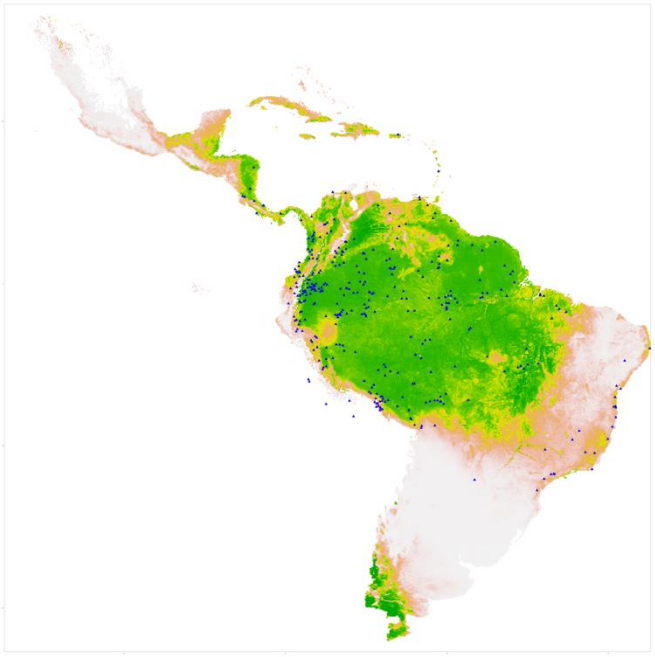


Pouteria bilocularis

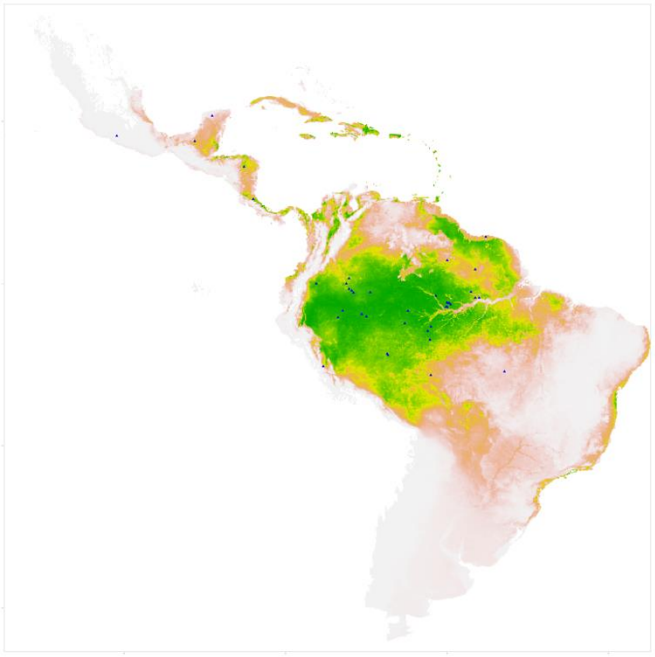
Pouteria bracteata



Pouteria buenaventurensis

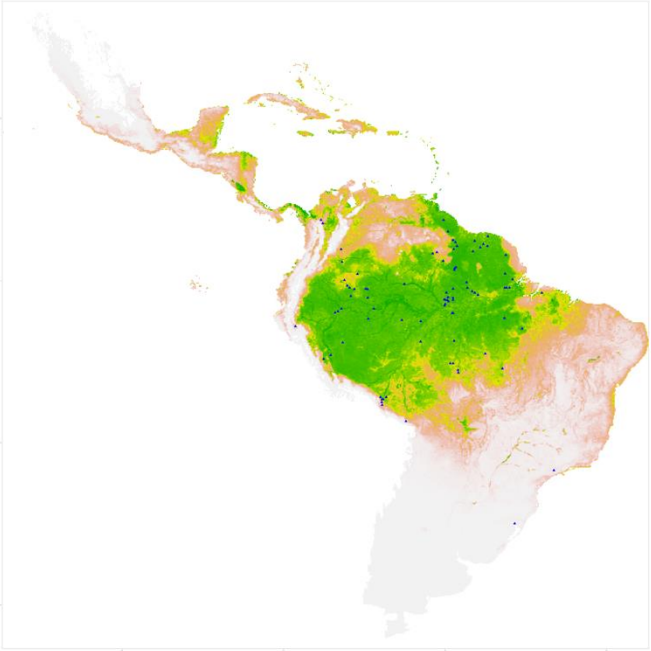


Pouteria caimito

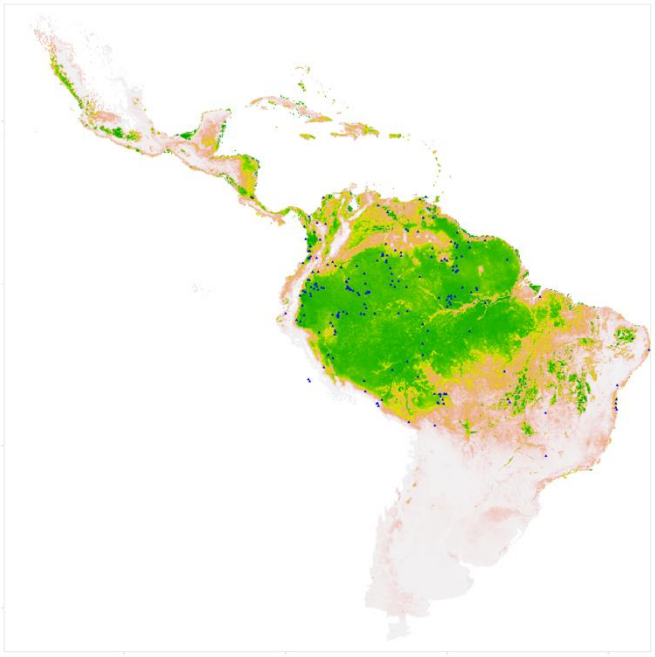


Pouteria campanulata

Pouteria campechiana

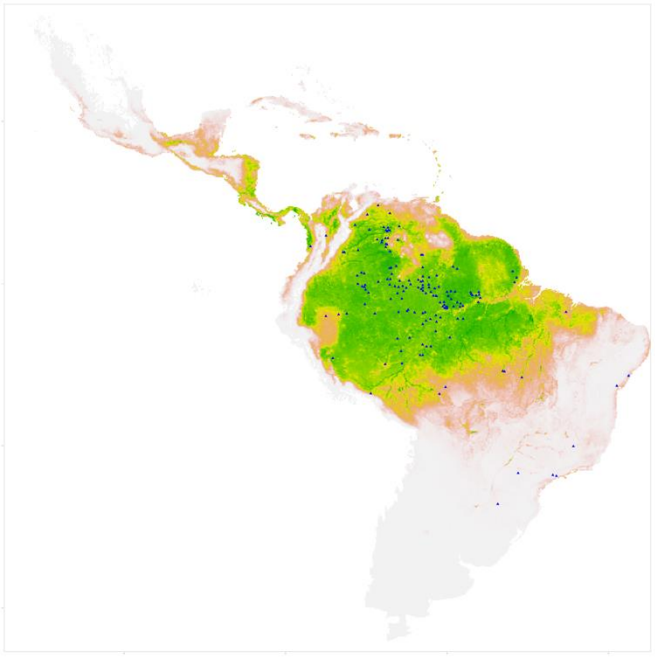


Pouteria cladantha

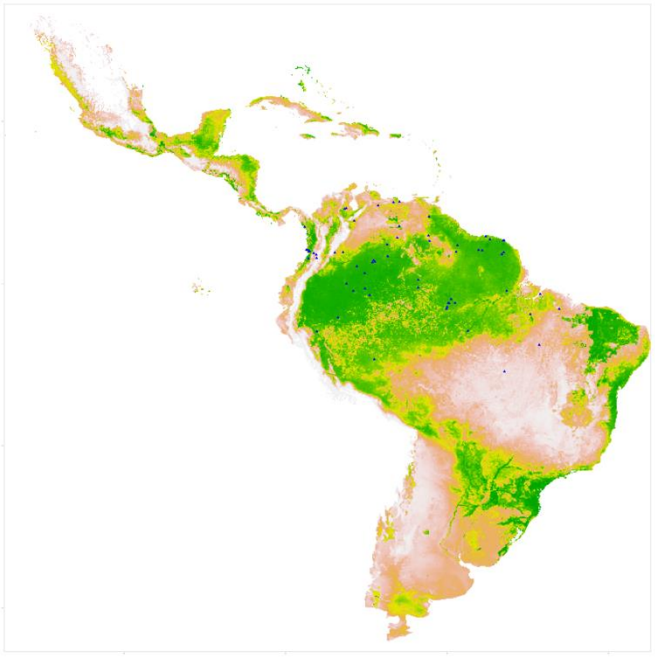


Pouteria cuspidata

Pouteria egregia

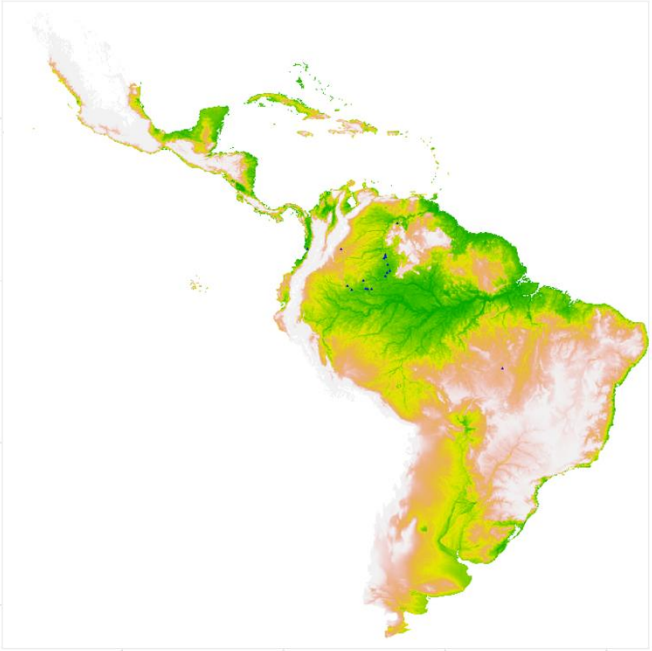


Pouteria elegans

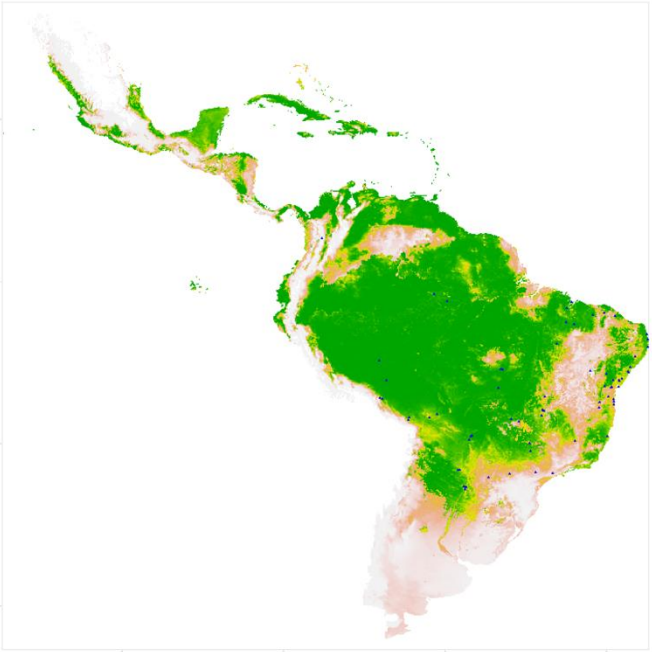


Pouteria eugeniifolia

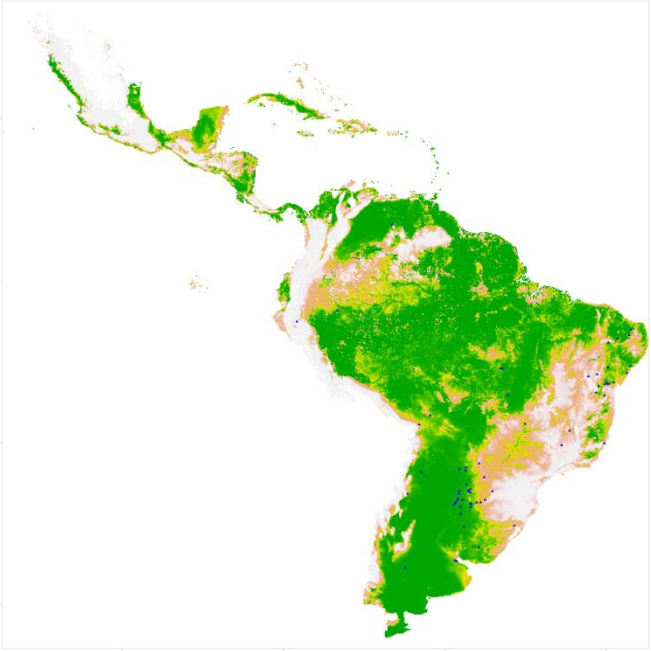
Pouteria franciscana



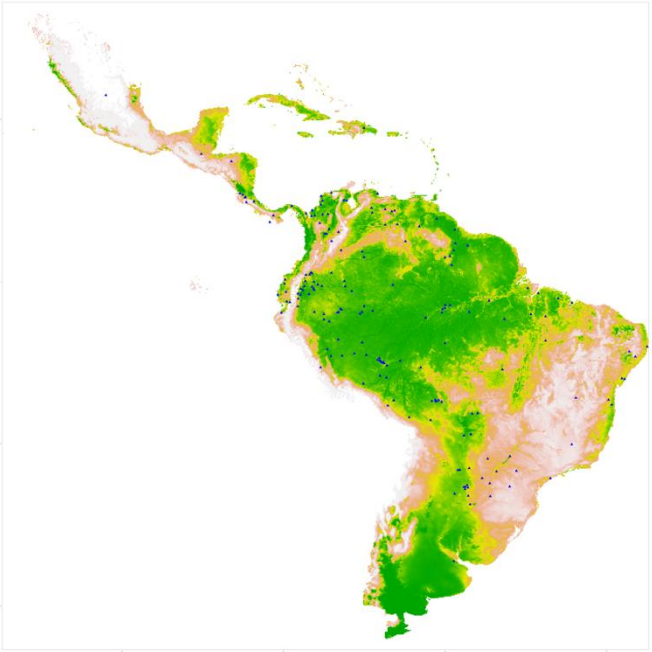
Pouteria gabrielensis



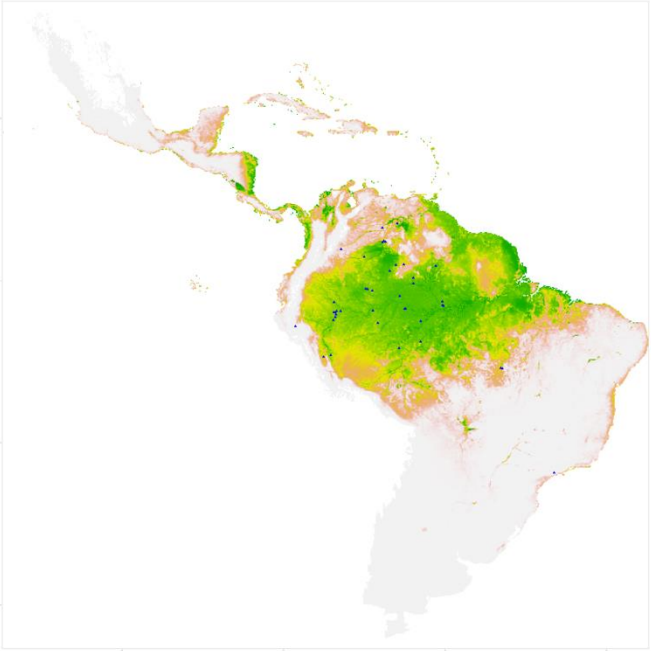
Pouteria gardneri



Pouteria gardneriana

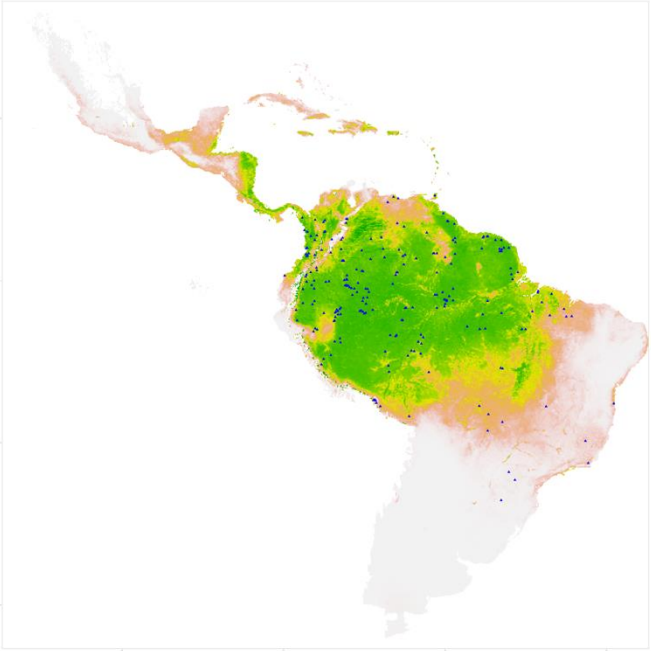


Pouteria glomerata



Pouteria gomphiifolia

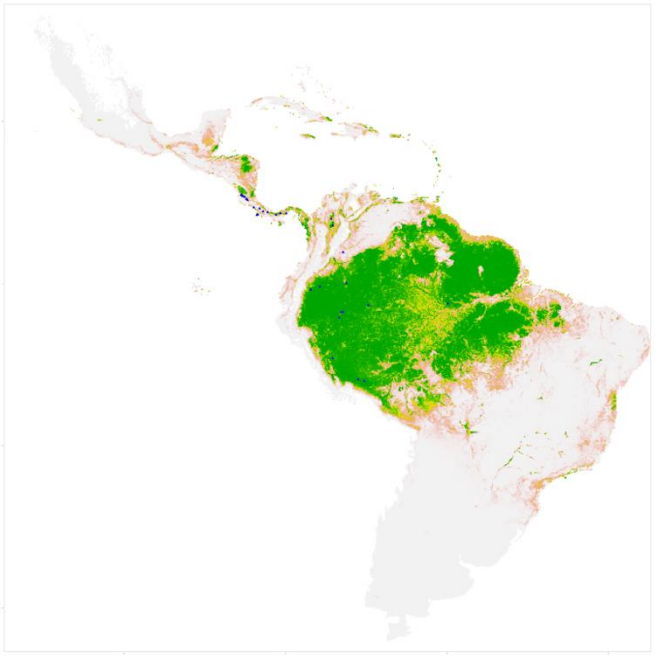
Pouteria grandiflora



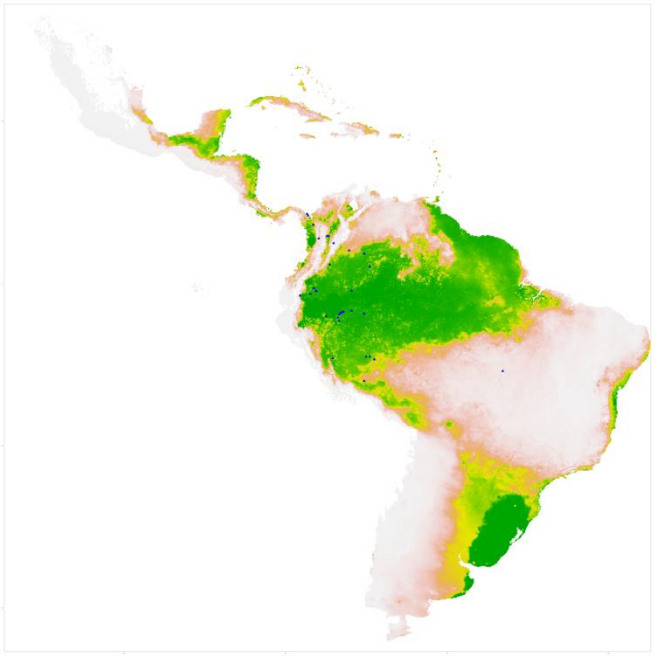
Pouteria guianensis

Pouteria hispida

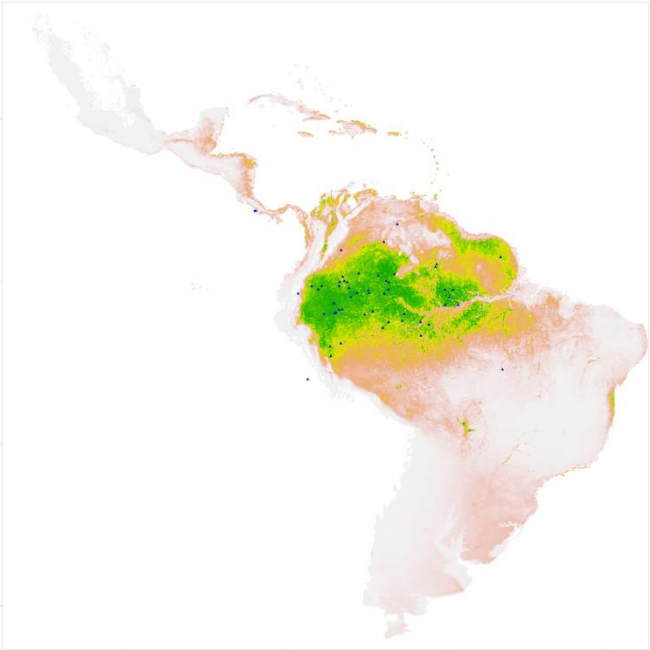
Pouteria izabalensis



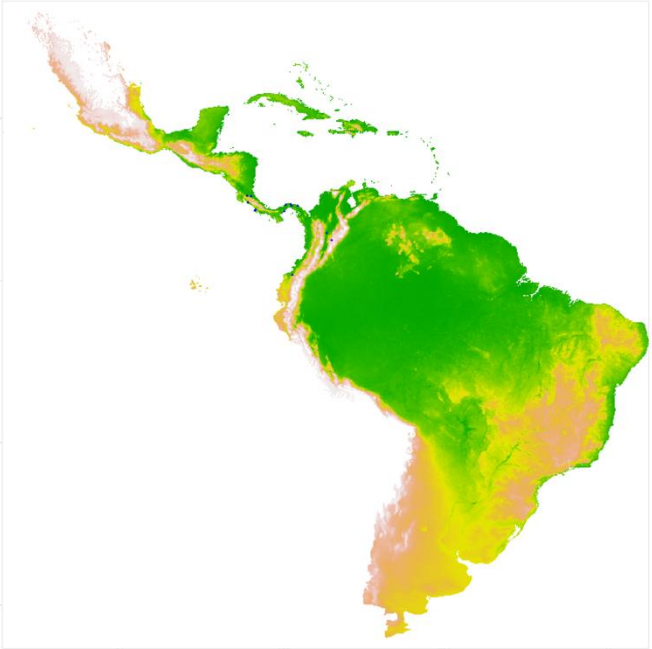
Pouteria juruana



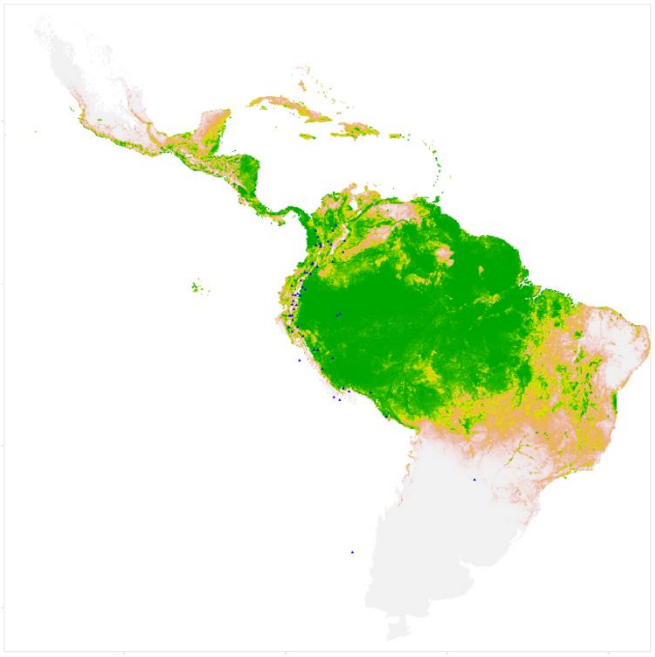
Pouteria krukovii



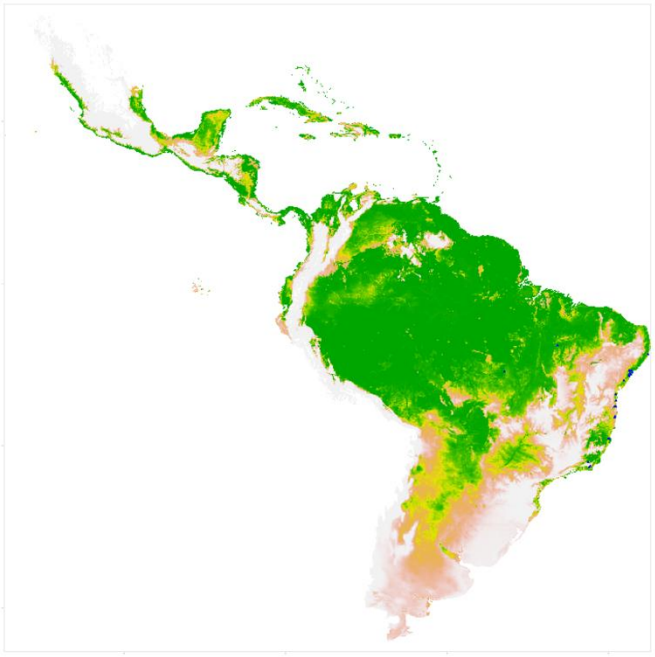
Pouteria laevigata



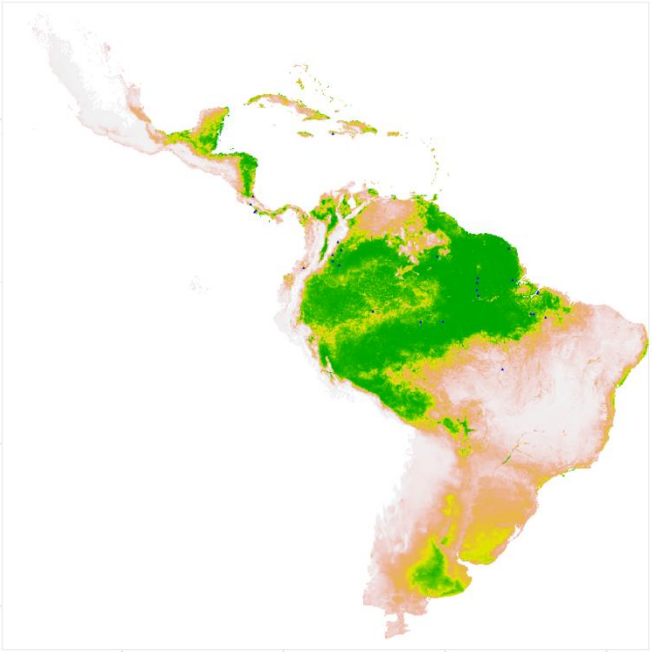
Pouteria leptopedicellata



Pouteria lucuma



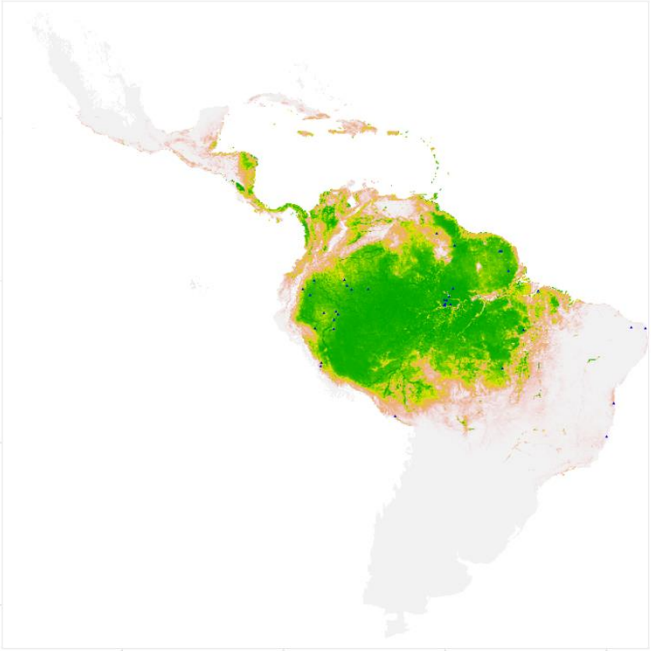
Pouteria macahensis



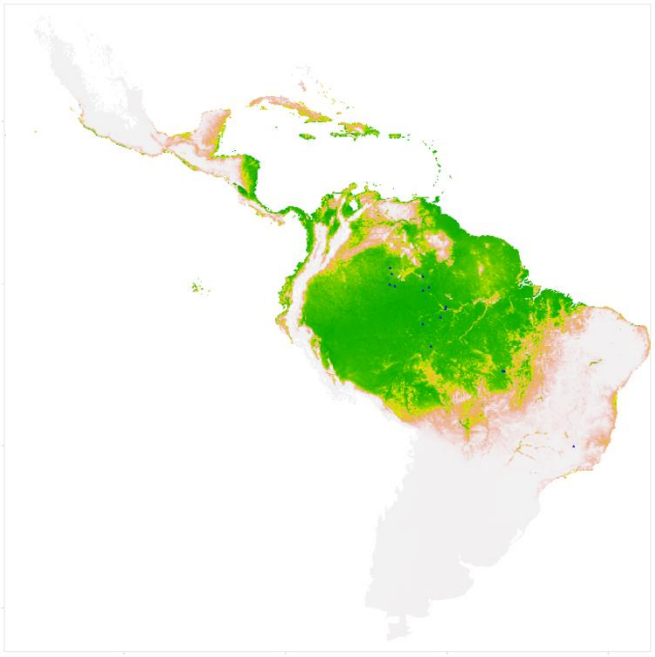
Pouteria macrocarpa

Pouteria macrophylla

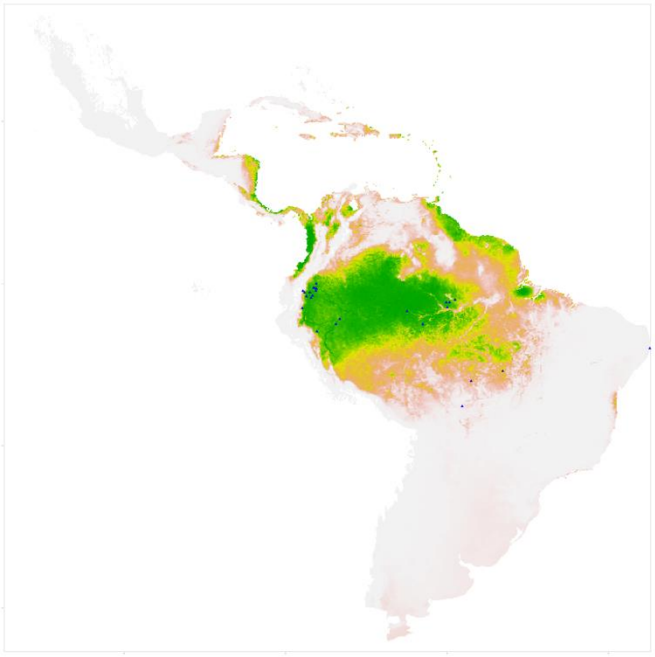
Pouteria multiflora



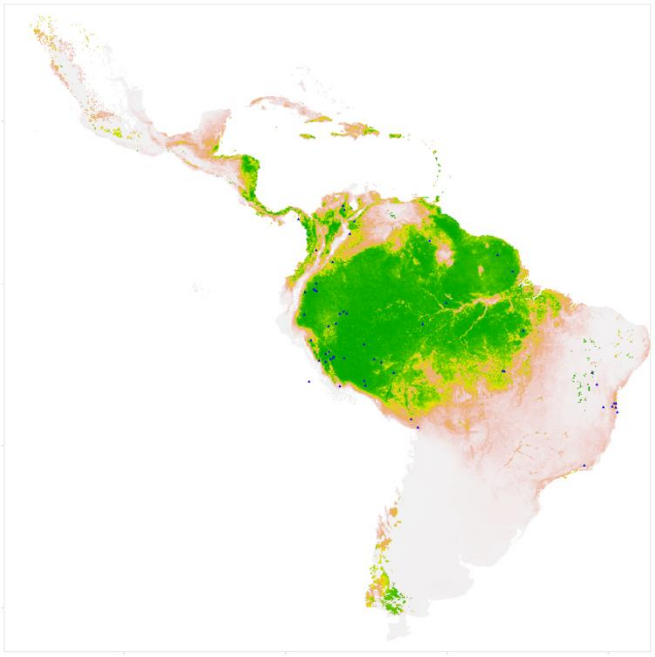
Pouteria oblanceolata



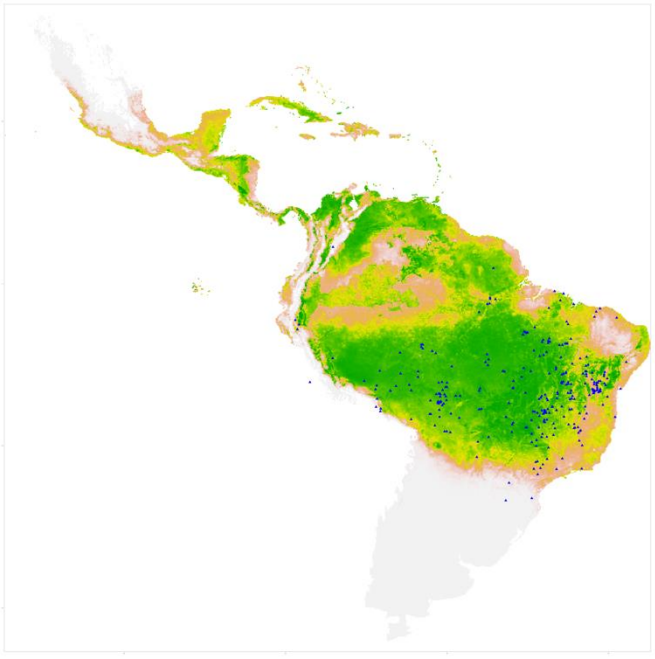
Pouteria pachyphylla



Pouteria platyphylla

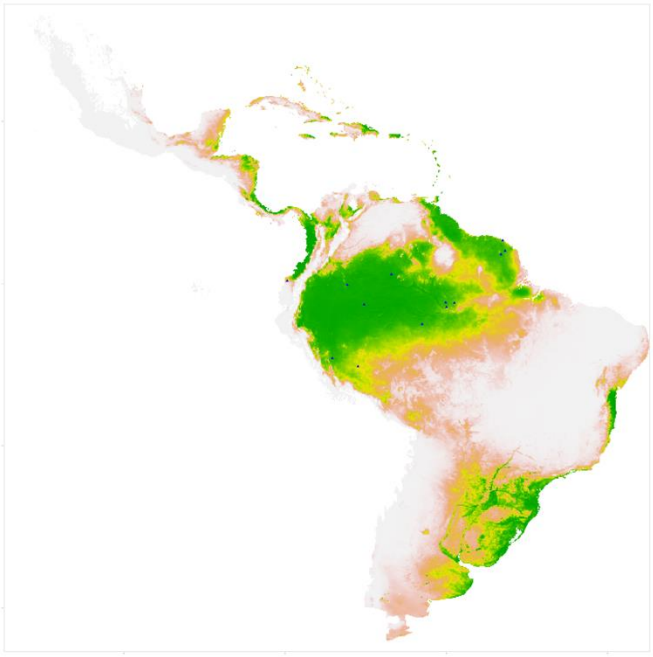


Pouteria procera

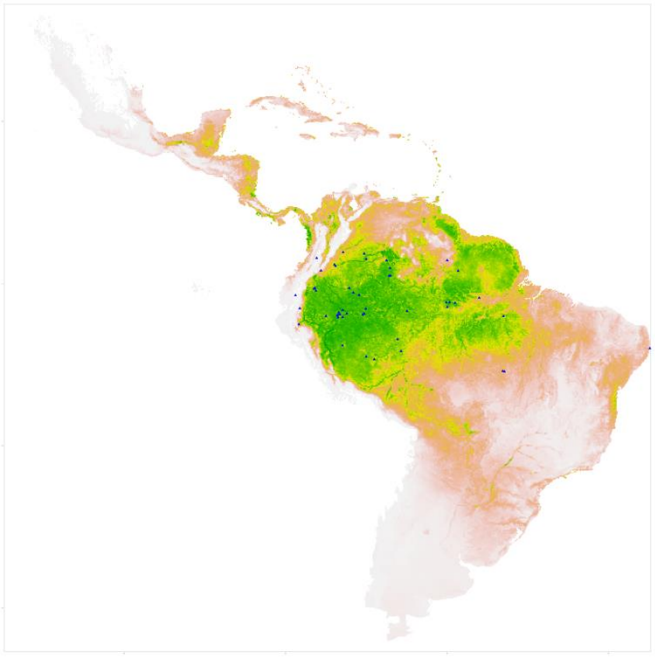


Pouteria ramiflora

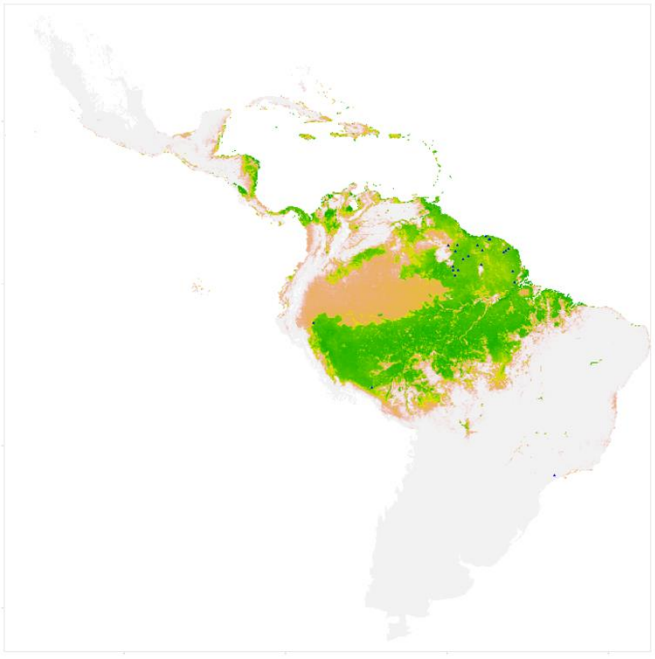
Pouteria reticulata



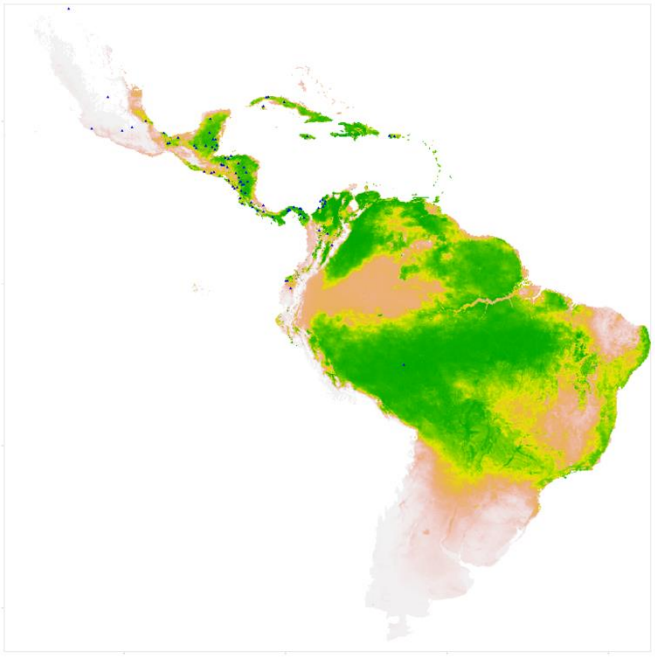
Pouteria retinervis



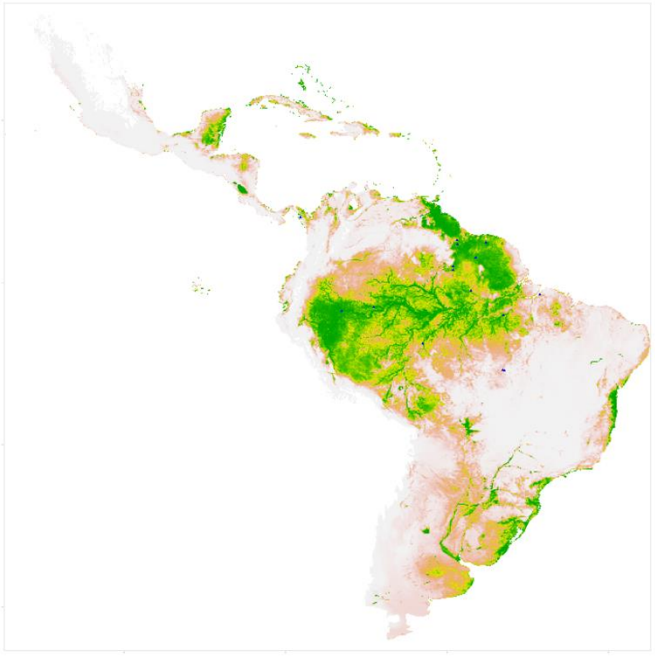
Pouteria rostrata



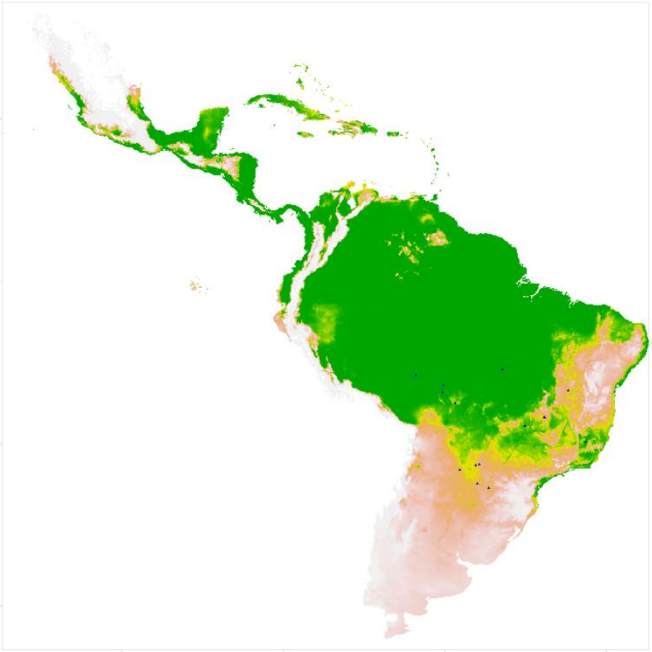
Pouteria sagotiana



Pouteria sapota



Pouteria speciosa

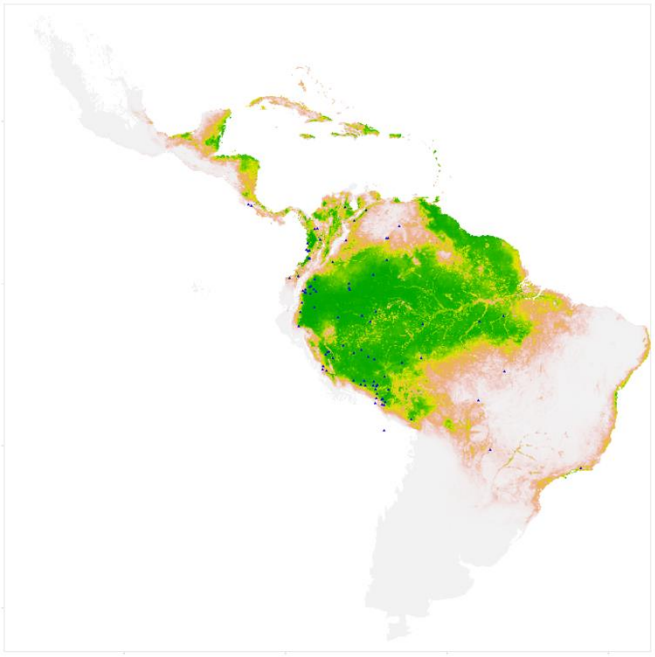


Pouteria subcaerulea

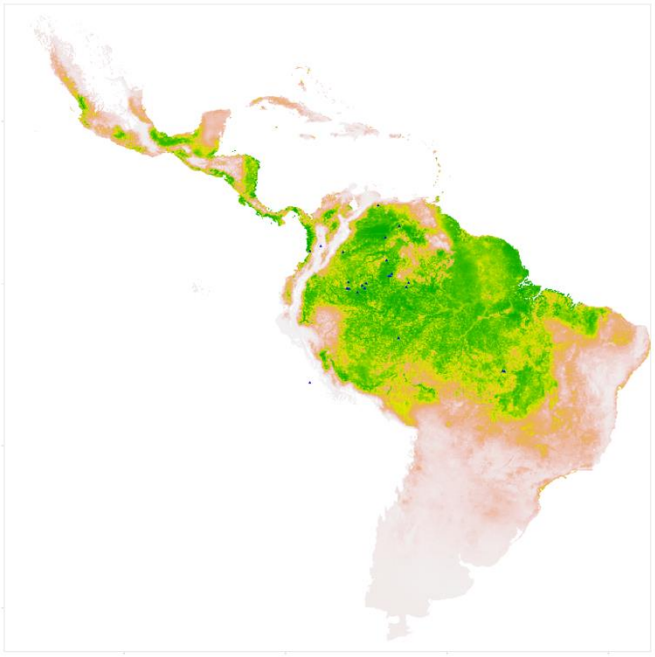
Pouteria subrotata

Pouteria surumuensis

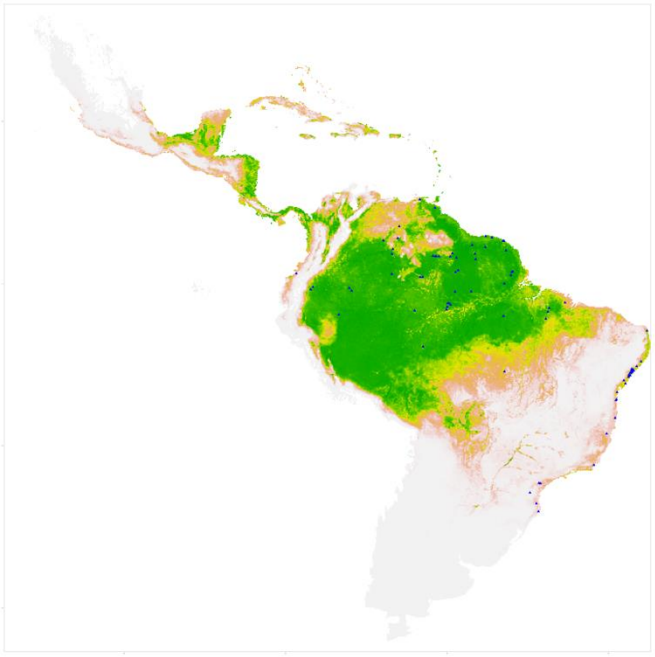
Pouteria trigonosperma



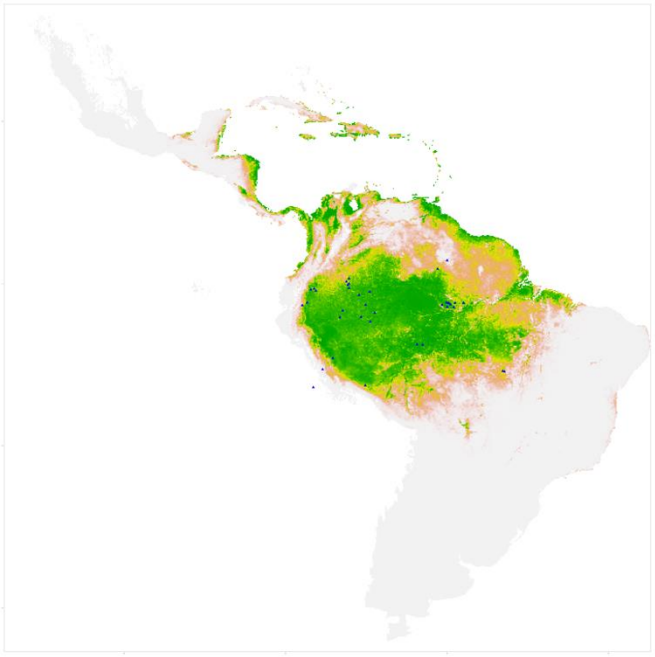
Pouteria trilocularis



Pouteria ucuqui



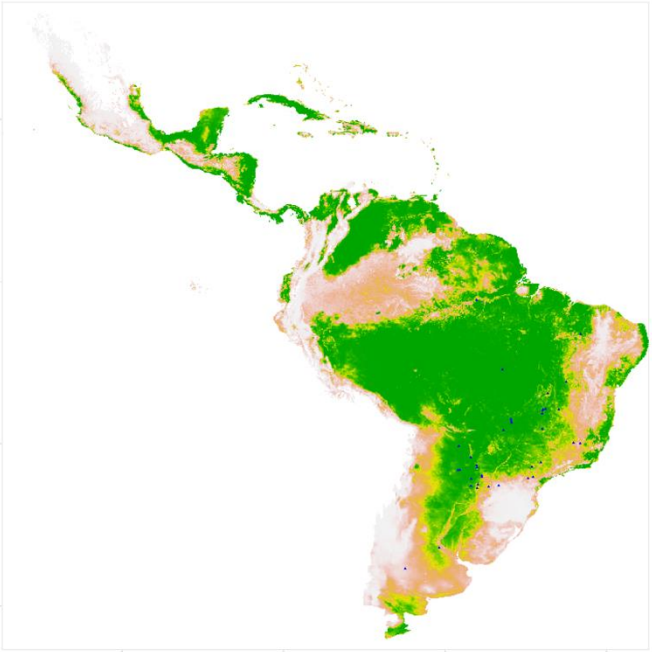
Pouteria venosa



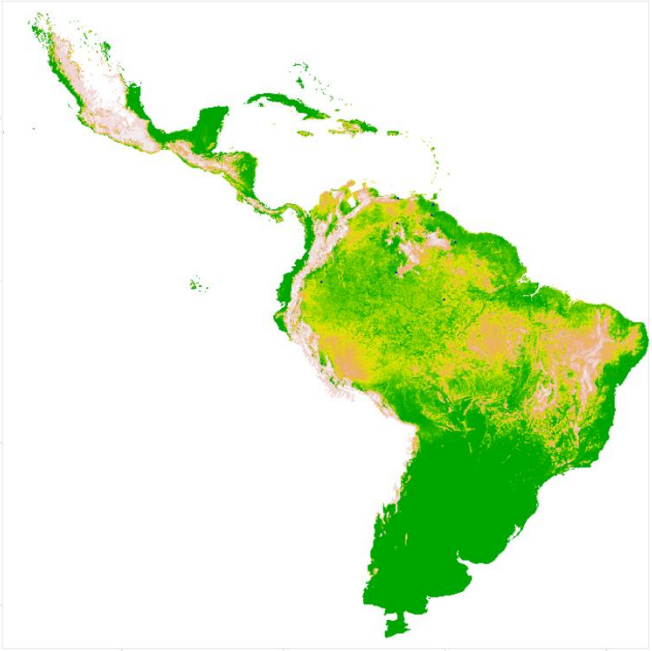
Pouteria vernicosa

Pouteria virescens

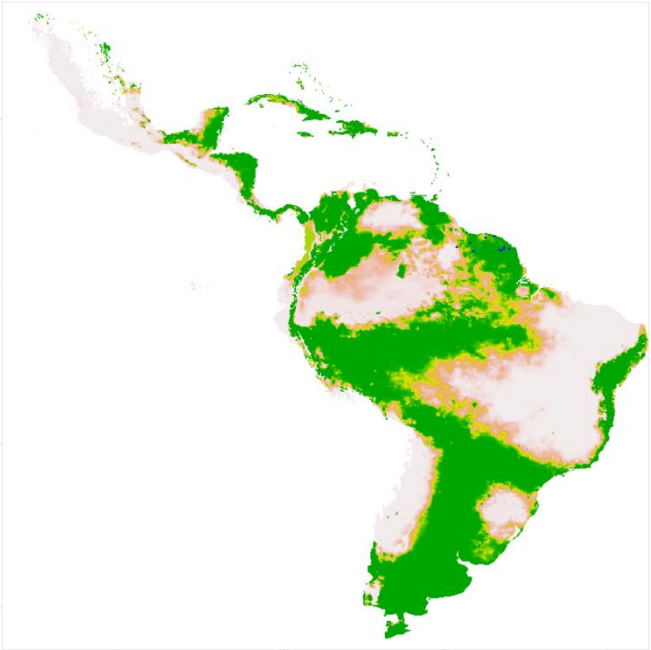
Pradosia beardii



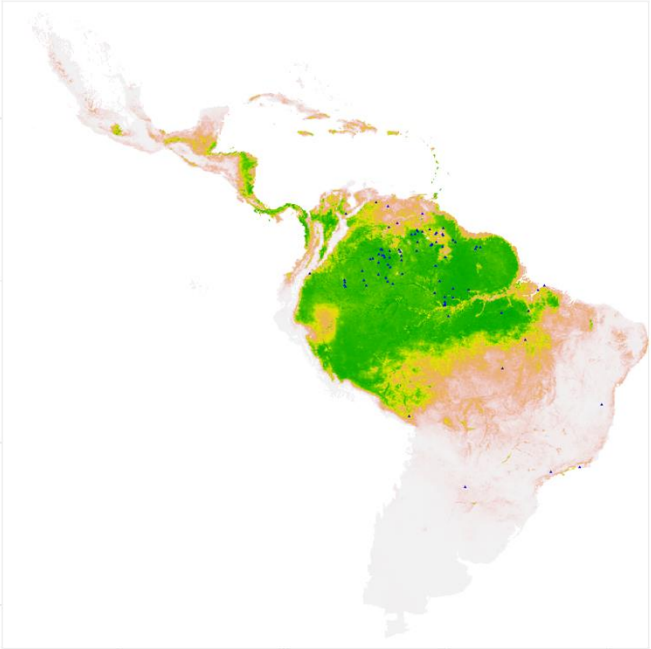
Pradosia brevipes



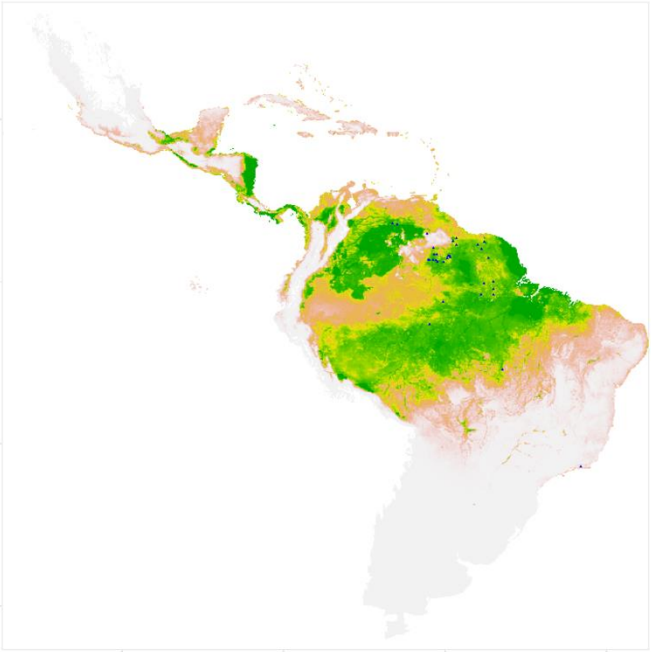
Pradosia grisebachii



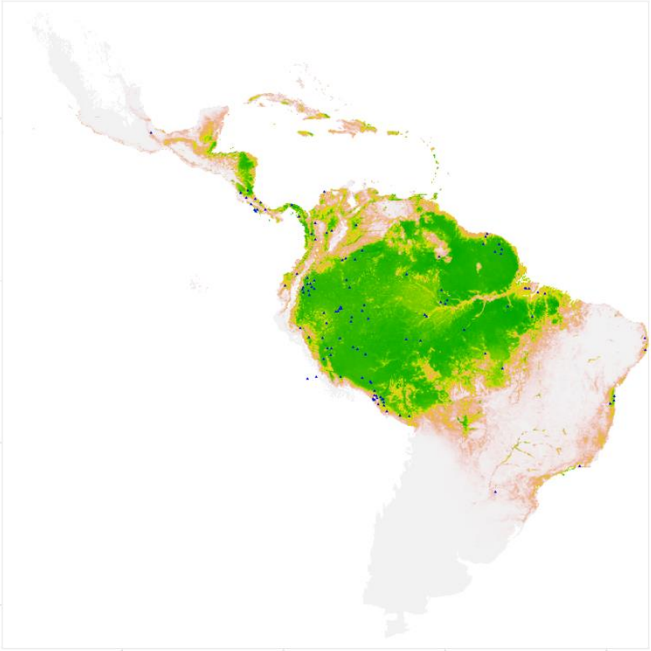
Pradosia ptychandra



Pradosia schomburgkiana

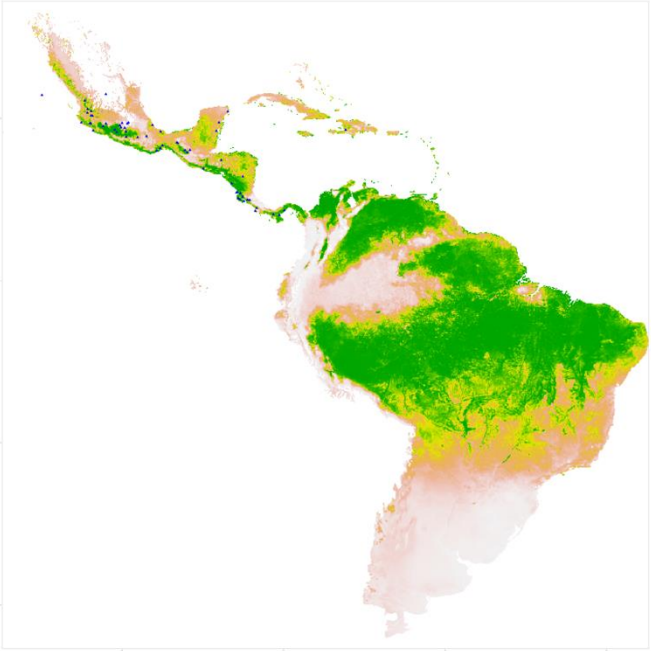


Pradosia surinamensis

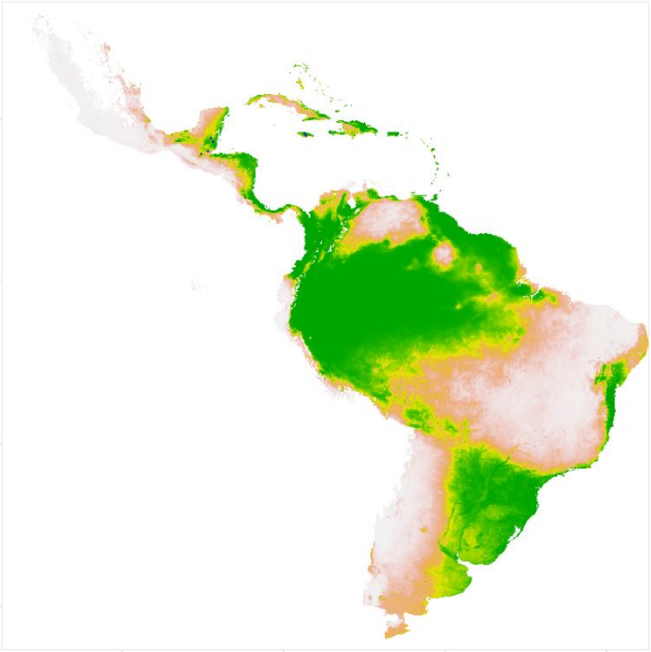


Sarcaulus brasiliensis

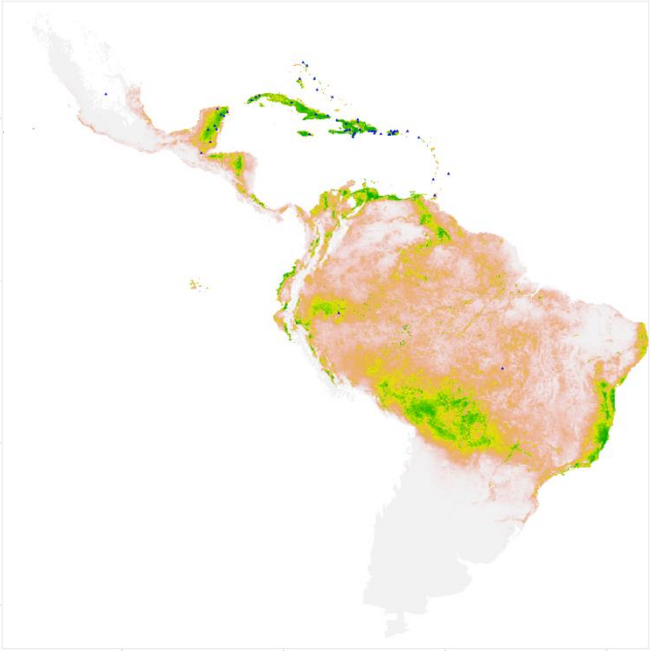
Sarcaulus inflexus



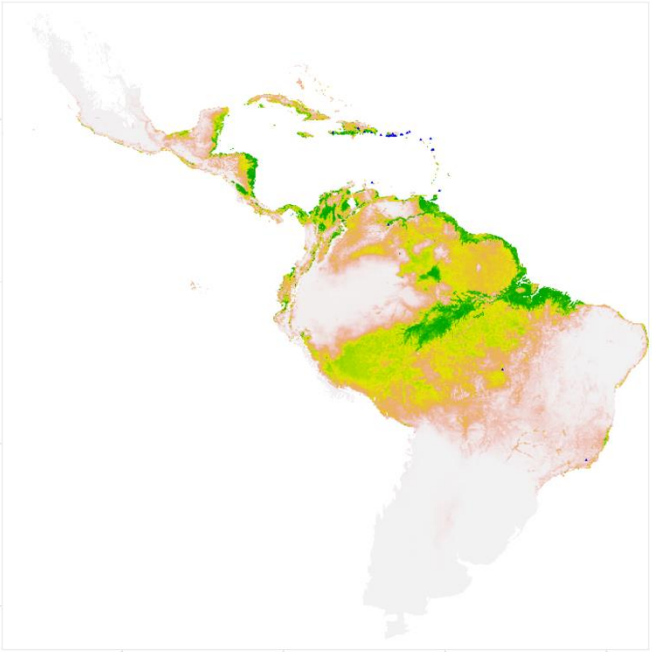
Sideroxylon capiri



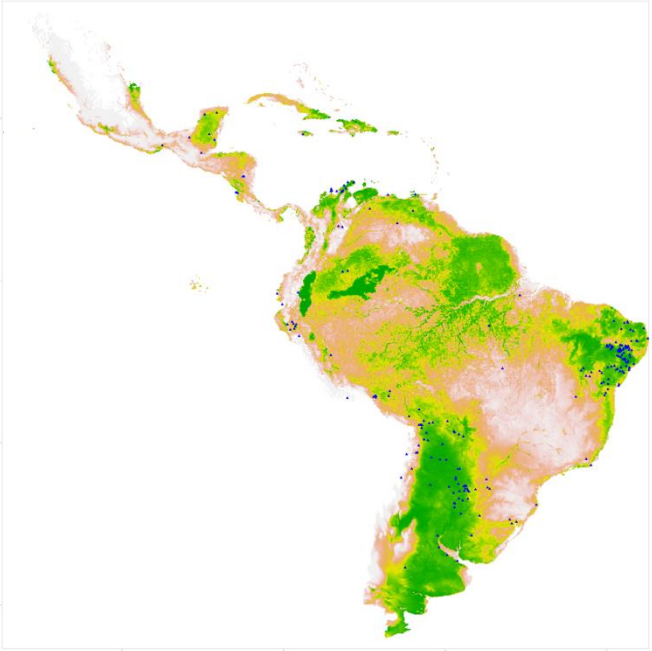
Sideroxylon floribundum



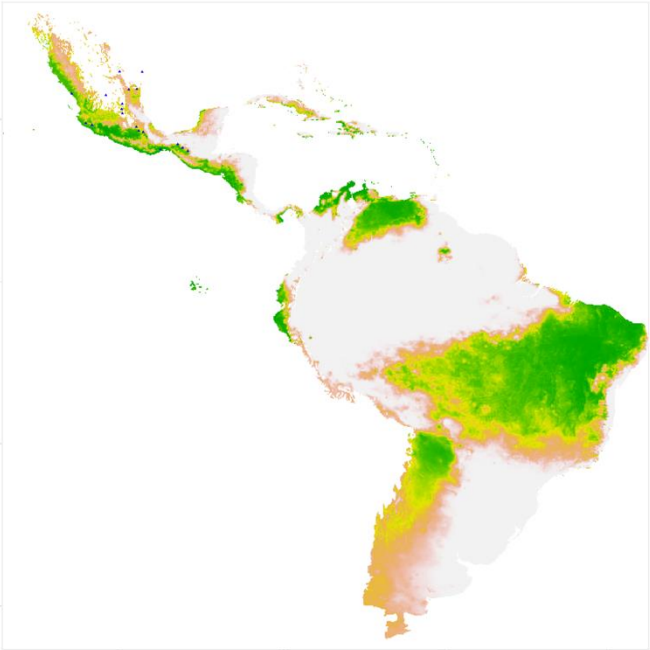
Sideroxylon foetidissimum



Sideroxylon obovatum

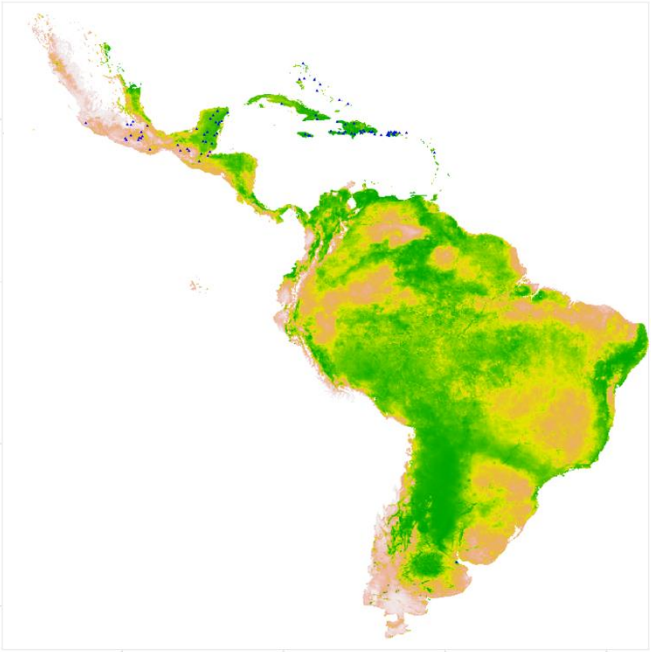


Sideroxylon obtusifolium

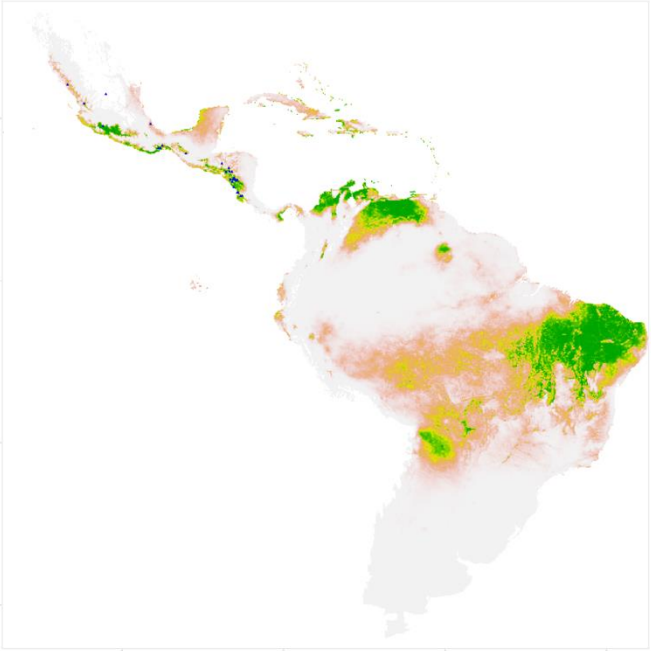


Sideroxylon palmeri

Sideroxylon portoricense



Sideroxylon salicifolium



Sideroxylon stenospermum

Appendix 3.1. Species Distribution Models.

Appendix 3.2

ID (Fig. 2)	District	Region of Colombia	Mean number of species per pixel across each District	Number of unique species per District
1	Macuira	Caribbean	23	48
2	Alta Guajira	Caribbean	22	48
3	Baja Guajira y Alto Cesar	Caribbean	43	45
4	Marocaso	Caribbean	57	89
5	Guachaca	Caribbean	50	114
6	Santa Marta Enclaves Azonales	Caribbean	52	89
7	Delta del Magdalena	Caribbean	56	61
8	Chundua	Caribbean	21	84
9	Maria y Piojo	Caribbean	80	111
10	Aracataca	Caribbean	57	84
11	Cartagena	Caribbean	63	114
12	Caracolicito	Caribbean	59	82
13	Ariguani-Cesar	Caribbean	62	114
14	Acandi-San Blas	Choco	84	113
15	Turbo	Choco	85	113
16	Sinu-San Jorge	Choco/Caribbean	77	118
17	La Gloria	Caribbean	71	113
18	Catatumbo	Catatumbo	77	115
19	Tacarcuna	Choco	59	105
20	Aspave-El Limon-Pirre	Choco	91	115
21	Jurado	Choco	93	113
22	Rio Sucio	Choco	75	113
23	Lebrija	Magdalena valley	93	117
24	Murri	Choco	79	113
25	Nechi	Magdalena valley/Caribbean	73	116
26	Baudo	Choco	83	102
27	Carare	Magdalena valley	77	116
28	Perija	Andes	29	89
29	Arauca-Apure	Llanos	70	108

ID (Fig. 2)	District	Region of Colombia	Mean number of species per pixel across each District	Number of unique species per District
30	Piedemonte Casanare-Arauca	Llanos	61	113
31	Utria	Choco	87	110
32	Alto Atrato-San Juan	Choco	72	87
33	Casanare	Llanos	78	116
34	Maipures	Llanos	52	110
35	Sabanas Altas	Llanos	66	116
36	Piedemonte Meta	Llanos	55	102
37	Selvas del Norte de Guaviare	Amazon	103	115
38	Tumaco	Choco	57	101
39	Micay	Choco	53	101
40	Macarena	Macarena	82	109
41	Ariari-Guayabero	Macarena	80	115
42	Barbacoas	Choco	42	91
43	Florencia	Amazon	68	113
44	Kofan	Amazon	81	95
45	Alto Putumayo	Amazon	91	112
46	Caguan	Amazon	93	115
47	Yari-Mariti	Amazon/Macarena	94	116
48	Complejo Vaupes	Amazon	98	116
50	Huitoto	Amazon	99	110
51	Ticuna	Amazon	100	111

Appendix 3.2. Species richness within Corzo's (2008) districts.
Species richness was measured as the maximum number of potential Sapotaceae species per pixel within each district by summarising thresholded species distribution models of Sapotaceae and intercepting that summary with the biogeographic districts defined by Corzo (2008).